

Davis, D.  
09/17/05 Page 1  
Seq ID 1 w/Intef

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: October 10, 2003, 17:22:39 ; Search time 406 Seconds  
(without alignments)  
40.341 Million cell updates/sec

Title: US-09-991-795-1

Perfect score: 91

Sequence: 1 KSPDQOETVDGNLIRY 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 5728757 seqs, 909918778 residues

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending Patents, AA Main:\*

1: /cgn2\_6/ptodata/1/paa/US06\_PCTUS\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US080\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US081\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep.\*  
8: /cgn2\_6/ptodata/1/paa/US084\_COMB.pep.\*  
9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep.\*  
10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*  
11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*  
12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*  
13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep.\*  
14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*  
15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*  
16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*  
17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*  
18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*  
19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*  
20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep.\*  
21: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep.\*  
22: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*  
23: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep.\*  
24: /cgn2\_6/ptodata/1/paa/US09A\_COMB.pep.\*  
25: /cgn2\_6/ptodata/1/paa/US09B\_COMB.pep.\*  
26: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*  
27: /cgn2\_6/ptodata/1/paa/US101\_COMB.pep.\*  
28: /cgn2\_6/ptodata/1/paa/US102\_COMB.pep.\*  
29: /cgn2\_6/ptodata/1/paa/US103\_COMB.pep.\*  
30: /cgn2\_6/ptodata/1/paa/US104\_COMB.pep.\*  
31: /cgn2\_6/ptodata/1/paa/US105\_COMB.pep.\*  
32: /cgn2\_6/ptodata/1/paa/US106\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	91	100.0	18	25	US-09-992-067-1

Result No.	Score	Query Match	Length DB	ID	Description
2	91	100.0	18	25	US-09-992-067A-1
3	91	100.0	339	28	PCT-US01-08656-10252
4	91	100.0	339	28	US-10-273-573-10252
5	91	100.0	648	1	PCT-US03-09774-5
6	91	100.0	737	28	US-10-221-279-9937
7	91	100.0	838	32	US-60-230-445-1056
8	91	100.0	847	1	PCT-US01-14827-14550
9	91	100.0	900	1	PCT-US02-29221-2
10	91	100.0	930	22	PCT-US03-15711-126
11	91	100.0	930	22	US-09-791-537-12639
12	91	100.0	930	22	US-09-791-537-53361
13	91	100.0	930	22	US-09-791-537-118834
14	91	100.0	930	30	US-10-440-464-126
15	91	100.0	930	30	US-10-444-575-6
16	91	100.0	971	1	PCT-US01-14827-14552
17	86	94.5	123	28	PCT-US01-08656-10251
18	86	94.5	123	28	US-10-273-573-10251
19	80.5	88.5	19	25	US-09-993-295-2
20	79	86.8	16	23	PCT-US02-31642-356
21	79	86.8	16	23	US-09-826-290-44
22	79	86.8	16	28	US-10-244-715A-356
23	79	86.8	16	28	US-10-264-309-356
24	79	86.8	16	31	US-10-624-429-281
25	57	62.6	932	30	US-10-444-575-4
26	50	54.9	16	30	US-10-444-575-1
27	47.5	52.2	853	1	PCT-US02-38526-513
28	47.5	52.2	853	22	US-09-791-537-83086
29	47.5	52.2	853	22	US-09-791-537-151450
30	47.5	52.2	853	32	US-60-384-450-29
31	47.5	52.2	858	22	US-09-791-537-10439
32	47.5	52.2	858	22	US-09-791-537-116884
33	47.5	52.2	858	28	US-10-219-051B-9268
34	47.5	52.2	859	1	PCT-US02-38437-35
35	46	49.5	601	27	US-10-156-761-14651
36	45	49.5	114	30	US-10-424-599-177890
37	45	49.5	428	21	US-09-708-427-6340
38	45	49.5	429	21	US-09-708-427-6339
39	45	49.5	431	21	US-09-708-427-6338
40	44	48.4	81	30	US-10-437-962-129305
41	44	48.4	86	22	US-09-791-537-107293
42	43	47.3	432	27	US-10-156-761-14180
43	43	47.3	432	30	US-10-424-599-216174
44	43	47.3	626	28	US-10-238-655-538
45	43	47.3	640	28	US-10-219-999-35481

## ALIGNMENTS

RESULT 1  
US-09-992-067-1  
; Sequence 1, Application US/09992067  
; GENERAL INFORMATION:  
; APPLICANT: Jackowski, George  
; TITLE OF INVENTION: Intracellular Trypsin Inhibitor Biopolymer Markers Indicative of Ins  
; FILE REFERENCE: 2132.100  
; CURRENT APPLICATION NUMBER: US/09/992,067  
; CURRENT FILING DATE: 2001-11-21  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-992-067-1

Query Match 100.0%; Score 91; DB 25; Length 18;  
Best Local Similarity 100.0%; Pred. No. 2.7e-08;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1 KSPDQOETVDGNLIRY 18  
|||||

Db 1 KSPSEQETVLDGNLIIRY 18

## RESULT 2

US-09-992-067A-1  
Sequence 1, Application US/09992067A

GENERAL INFORMATION:

APPLICANT: Jackowski, George

TITLE OF INVENTION: Interleukin Typein Inhibitor Polypolymer Markers Indicative of In

TITLE OF INVENTION: Resistance

FILE REFERENCE: 2132.100

CURRENT APPLICATION NUMBER: US/09/992,067A

CURRENT FILING DATE: 2001-11-23

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn version 3.1

SEQ ID NO 1

LENGTH: 18

TYPE: PRT

ORGANISM: Homo sapiens

US-09-992-067A-1

Query Match 100.0%; Score 91; DB 25; Length 18;  
Best Local Similarity 100.0%; Pred. No. 2,7e-08;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KSPSEQETVLDGNLIIRY 18

Db 1 KSPSEQETVLDGNLIIRY 18

## RESULT 3

PCT-US01-08656-10252  
Sequence 10252, Application PC/TUS0108656

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc

TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES

FILE REFERENCE: 21272-066

CURRENT APPLICATION NUMBER: PCT/US01/08656

CURRENT FILING DATE: 2001-04-16

PRIOR APPLICATION NUMBER: 09/522,929

PRIOR FILING DATE: 2000-04-18

PRIOR APPLICATION NUMBER: 09/770,160

NUMBER OF SEQ ID NOS: 10994

SOFTWARE: Custom

SEQ ID NO 10252

LENGTH: 339

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: DOMAIN

LOCATION: (156)..(209)

OTHER INFORMATION: INHIBITOR HEAVY CHAIN CHANNEL IN domain identified by

OTHER INFORMATION: EMATRIX, accession number PD01101B, P-value=1.000e-40, raw score

OTHER INFORMATION: 21.53

NAME/KEY: DOMAIN

LOCATION: (174)..(339)

OTHER INFORMATION: von Willebrand factor type A domain identified by Pfam,

OTHER INFORMATION: accession name vwa, E-value=0.00046, Pfam score of 5.5

PCT-US01-08656-10252

Query Match 100.0%; Score 91; DB 1; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KSPSEQETVLDGNLIIRY 18

Db 124 KSPSEQETVLDGNLIIRY 141

## RESULT 4

US-10-273-10252  
Sequence 10252, Application US/10273573

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc

TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES

FILE REFERENCE: 21272-066

CURRENT APPLICATION NUMBER: US/10/273,573

CURRENT FILING DATE: 2002-10-18

PRIOR APPLICATION NUMBER: 09/522,929

PRIOR FILING DATE: 2000-04-18

PRIOR APPLICATION NUMBER: 09/770,160

PRIOR FILING DATE: 2001-01-26

NUMBER OF SEQ ID NOS: 10994

SOFTWARE: Custom

SEQ ID NO 10252

LENGTH: 339

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: DOMAIN

LOCATION: (156)..(209)

OTHER INFORMATION: INHIBITOR HEAVY CHAIN CHANNEL IN domain identified by

OTHER INFORMATION: EMATRIX, accession number PD01101B, P-value=1.000e-40, raw score

OTHER INFORMATION: 21.53

NAME/KEY: DOMAIN

LOCATION: (174)..(339)

OTHER INFORMATION: von Willebrand factor type A domain identified by Pfam,

OTHER INFORMATION: accession name vwa, E-value=0.00046, Pfam score of 5.5

US-10-273-573-10252

Query Match 100.0%; Score 91; DB 28; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KSPSEQETVLDGNLIIRY 18

Db 124 KSPSEQETVLDGNLIIRY 141

## RESULT 5

PCT-US01-09774-5  
Sequence 5, Application PC/TUS0309774

GENERAL INFORMATION:

APPLICANT: INCYTE CORPORATION; LAU, Preeti G.;

APPLICANT: TANG, Y. Tom; LEE, Soo Y.;

APPLICANT: GRIFFIN, Jennifer A.; ELIOTY, Vicki S.;

APPLICANT: KHARE, Reena; RAMKUMAR, Jayalaxmi;

APPLICANT: RICHARDSON, Thomas W.; ISON, Craig H.;

APPLICANT: HARALIA, April J.A.; KABLE, Amy E.;

APPLICANT: SWARNAKAR, Anita; CHANG, Hsin-Ru.;

APPLICANT: BHATIA, Umesh G.; BURRILL, John D.;

APPLICANT: LEE, Sally; BLAKE, Julie J.;

APPLICANT: HO, Anne; ZHENG, Wenjin;

APPLICANT: CHAWLA, Narinder K.; MARQUIS, Joseph P.;

APPLICANT: TRAN, Uyen K.; EMERLING, Brooke M.;

APPLICANT: MASON, Patricia M.; BECHA, Shanya D.;

APPLICANT: WARREN, Bridget A.; AU-YOUNG, Janice K.;

APPLICANT: LEE, Ernestine A.; GIERZEN, Kimberly J.;

APPLICANT: JIANG, Xin; JACKSON, Alan A.;

APPLICANT: CHANG, Hsin-Ru; BAUGHN, Mariah R.;

APPLICANT: WILSON, Amy D.; JIN, Pei;

APPLICANT: BULLOCH, Sean A.;

TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES

FILE REFERENCE: PF-1408 PCT

CURRENT APPLICATION NUMBER: PCT/US03/09774

CURRENT FILING DATE: 2003-03-26

PRIOR APPLICATION NUMBER: US 60/368,686

PRIOR FILING DATE: 2002-03-29

PRIOR APPLICATION NUMBER: US 60/378,205

PRIOR FILING DATE: 2002-05-03

PRIOR APPLICATION NUMBER: US 60/377,489

PRIOR FILING DATE: 2002-05-03

NUMBER OF SEQ ID NOS: 142

SOFTWARE: PERL Program

SEQ ID NO 5  
LENGTH: 648  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No: 7510116CD1  
PCT-US03-09774-5

Query Match 100.0%; Score 91; DB 1; Length 648;  
Best Local Similarity 100.0%; Pred. No. 2.9e-06;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSPQOETVLDGNLIIRY 18  
DB 224 KSPQOETVLDGNLIIRY 241

ND

RESULT 6  
US-10-221-279-9937  
Sequence 9937, Application US/10221279  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc  
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides  
FILE REFERENCE: 21272-046  
CURRENT APPLICATION NUMBER: US/10/221,279  
CURRENT FILING DATE: 2002-09-06  
PRIOR APPLICATION NUMBER: 09/574,454  
PRIOR FILING DATE: 2000-05-19  
PRIOR APPLICATION NUMBER: 09/519,705  
PRIOR FILING DATE: 2000-03-07  
NUMBER OF SEQ ID NOS: 12360  
SOFTWARE: Custom  
SEQ ID NO 9937  
LENGTH: 737  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-221-279-9937

Query Match 100.0%; Score 91; DB 28; Length 737;  
Best Local Similarity 100.0%; Pred. No. 3.4e-06;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSPQOETVLDGNLIIRY 18  
DB 31 KSPQOETVLDGNLIIRY 48

ND

RESULT 7  
US-60-230-445-1056  
Sequence 1056, Application US/60230445  
GENERAL INFORMATION:  
APPLICANT: Beasley, Ellen  
TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,  
TITLE OF INVENTION: AND USRS THEROF  
FILE REFERENCE: CL000765  
CURRENT APPLICATION NUMBER: US/60/230,445  
CURRENT FILING DATE: 2000-09-06  
NUMBER OF SEQ ID NOS: 3051  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1056  
LENGTH: 838  
TYPE: PRT  
ORGANISM: HUMAN  
US-60-230-445-1056

Query Match 100.0%; Score 91; DB 32; Length 838;  
Best Local Similarity 100.0%; Pred. No. 4e-06;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSPQOETVLDGNLIIRY 18

ND

DB 224 KSPQOETVLDGNLIIRY 241

RESULT 8  
PCT-US01-14827-14550  
Sequence 14550, Application PC/TUS0114827  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc  
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
FILE REFERENCE: 21272-104  
CURRENT APPLICATION NUMBER: PCT/US01/14827  
CURRENT FILING DATE: 2001-05-16  
PRIOR APPLICATION NUMBER: 09/577,408  
PRIOR FILING DATE: 2000-05-18  
NUMBER OF SEQ ID NOS: 16102  
SOFTWARE: Custom  
SEQ ID NO 14550  
LENGTH: 847  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: DOMAIN  
LOCATION: (256)..(309)  
OTHER INFORMATION: INHIBITOR HEAVY CHAIN CHANNEL IN domain identified by  
OTHER INFORMATION: EMATRIX, accession number PD01101B, p-value=1.000e-40, raw score  
OTHER INFORMATION: von Willebrand factor type A domain identified by Pfam,  
OTHER INFORMATION: accession name vwa, E-value=1.1e-08, Pfam score of 42.3  
PCT-US01-14827-14550

Query Match 100.0%; Score 91; DB 1; Length 847;  
Best Local Similarity 100.0%; Pred. No. 4e-06;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSPQOETVLDGNLIIRY 18  
DB 224 KSPQOETVLDGNLIIRY 241

ND

RESULT 9  
PCT-US02-29221-2  
Sequence 2, Application PC/TUS0229221  
GENERAL INFORMATION:  
APPLICANT: INCYTE GENOMICS, INC.  
APPLICANT: SPRAGUE, William W.  
APPLICANT: CHAMLA, Nardger K.  
APPLICANT: WARREN, Bridget A.  
APPLICANT: TANG, Y. Tom  
APPLICANT: ELLIOTT, Vicki S.  
APPLICANT: MARQUIS, Joseph P.  
APPLICANT: LI, Joana X.  
APPLICANT: GRIFPIN, Jennifer A.  
APPLICANT: GIETZEN, Kimberly J.  
APPLICANT: YANG, Junming  
APPLICANT: LU, Dying Aina M.  
APPLICANT: EMERLING, Brooke M.  
APPLICANT: DUGAN, Brendan M.  
APPLICANT: RICHARDSON, Thomas W.  
APPLICANT: LEE, Soo Yeun  
APPLICANT: RAMKUMAR, Jayalaxmi  
APPLICANT: BECHA, Shanya D.  
APPLICANT: LEHR-MASON, Patricia M.  
APPLICANT: SWARNAKAR, Anita  
APPLICANT: TRAN, Uyen K.  
APPLICANT: KABLE, Amy E.  
APPLICANT: HAPALIA, April J.A.  
APPLICANT: KARE, Keena  
TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES  
FILE REFERENCE: PF-1186 PCT  
CURRENT APPLICATION NUMBER: PCT/US02/29221  
CURRENT FILING DATE: 2002-09-13

```

; PRIOR APPLICATION NUMBER: US 60/322,196
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: US 60/324,134
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/327,233
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: US 60/346,198
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 60/343,980
; PRIOR FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: US 60/348,887
; PRIOR FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: US 60/332,423
; PRIOR FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: US 60/334,145
; PRIOR FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: US 60/334,229
; PRIOR FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: US 60/337,451
; PRIOR FILING DATE: 2001-12-06
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 900
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No: 7500515CD1
PCT-US02-29221-2
```

```
Query Match      100.0%; Score 91; DB 1; Length 900;
Best Local Similarity 100.0%; Pred. No. 4.4e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 KSPSEQETVLDGNLIIRY 18
Db      224 KSPSEQETVLDGNLIIRY 241
```

ND

```

RESULT 10
PCT-US03-15711-126
; Sequence 126, Application PC/TUS0315711
; GENERAL INFORMATION:
; APPLICANT: SUGEN, INC.
; TITLE OF INVENTION: NOVEL BIOMARKERS OF TYROSINE KINASE INHIBITOR EXPOSURE
; FILE REFERENCE: 038602/1593
; CURRENT APPLICATION NUMBER: PCT/US03/15711
; CURRENT FILING DATE: 2003-05-19
; PRIOR APPLICATION NUMBER: 60/380,872
; PRIOR FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/448,922
; PRIOR FILING DATE: 2003-02-24
; PRIOR APPLICATION NUMBER: 60/448,874
; PRIOR FILING DATE: 2003-02-24
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 126
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US03-15711-126
```

```
Query Match      100.0%; Score 91; DB 1; Length 930;
Best Local Similarity 100.0%; Pred. No. 4.6e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 KSPSEQETVLDGNLIIRY 18
Db      224 KSPSEQETVLDGNLIIRY 241
```

ND

```

RESULT 11
US-09-791-537-12639
; Sequence 12639, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBER
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12639
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-537-12639
```

```
Query Match      100.0%; Score 91; DB 22; Length 930;
Best Local Similarity 100.0%; Pred. No. 4.6e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 KSPSEQETVLDGNLIIRY 18
Db      224 KSPSEQETVLDGNLIIRY 241
```

ND

```

RESULT 12
US-09-791-537-53361
; Sequence 53361, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBER
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 53361
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-791-537-53361
```

```
Query Match      100.0%; Score 91; DB 22; Length 930;
Best Local Similarity 100.0%; Pred. No. 4.6e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 KSPSEQETVLDGNLIIRY 18
Db      224 KSPSEQETVLDGNLIIRY 241
```

ND

```

RESULT 13
US-09-791-537-118834
; Sequence 118834, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBER
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
```



SEQ ID NO 118834  
LENGTH: 930  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-791-537-118834

Query Match 100.0%; Score 91; DB 22; Length 930;  
Best Local Similarity 100.0%; Pred. No. 4.6e-06;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSPROQETVVDGNLIIRY 18  
DB 224 KSPROQETVVDGNLIIRY 241

RESULT 14  
US-10-440-464-126

Sequence 126, Application US/10440464  
GENERAL INFORMATION:  
APPLICANT: DEPRIMO, SAMUEL  
APPLICANT: O'FARRELL, ANNE-MARIE  
APPLICANT: MORIMOTO, ALYSSA  
APPLICANT: SMOLICH, BEVERLY  
APPLICANT: MANNING, WILLIAM  
APPLICANT: WALTER, SARAH  
APPLICANT: CHERINGTON, JULIE  
APPLICANT: SCHILLING, JIM  
TITLE OF INVENTION: NOVEL BIOMARKERS OF TYROSINE KINASE INHIBITOR EXPOSURE  
TITLE OF INVENTION: AND ACTIVITY IN MAMMALS  
FILE REFERENCE: 038602/1592  
CURRENT APPLICATION NUMBER: US/10/440,464  
CURRENT FILING DATE: 2003-05-19  
PRIOR APPLICATION NUMBER: 60/380,872  
PRIOR FILING DATE: 2002-05-17  
PRIOR APPLICATION NUMBER: 60/448,922  
PRIOR FILING DATE: 2003-02-24  
PRIOR APPLICATION NUMBER: 60/448,874  
PRIOR FILING DATE: 2003-02-24  
NUMBER OF SEQ ID NOS: 185  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 126  
LENGTH: 930  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-440-464-126

Query Match 100.0%; Score 91; DB 30; Length 930;  
Best Local Similarity 100.0%; Pred. No. 4.6e-06;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSPROQETVVDGNLIIRY 18  
DB 224 KSPROQETVVDGNLIIRY 241

RESULT 15  
US-10-444-575-6

Sequence 6, Application US/10444575  
GENERAL INFORMATION:  
APPLICANT: University of Connecticut Health Center  
APPLICANT: Kuchel, George A  
APPLICANT: Zhu, Qing  
TITLE OF INVENTION: Compositions and Methods Relating to Detrusor Estrogen-Regulated  
FILE REFERENCE: UCT-0035  
CURRENT APPLICATION NUMBER: US/10/444,575  
CURRENT FILING DATE: 2003-05-22  
PRIOR APPLICATION NUMBER: US 60/382,830  
PRIOR FILING DATE: 2002-05-23  
NUMBER OF SEQ ID NOS: 47  
SOFTWARE: Patentin version 3.2  
SEQ ID NO 6  
LENGTH: 930

TYPE: PRT  
ORGANISM: Homo sapiens  
PUBLICATION INFORMATION:  
AUTHORS: Tobe, T., Saguchi, K., Hashimoto, K., Miura, N.H., Tomita, M.,  
AUTHORS: Li, F., Wang, Y., Minoshima, S., and Shirizu, N.  
TITLE: Mapping of human inter-alpha-trypsin inhibitor family heavy  
TITLE: chain-related protein gene (ITIH1) to human chromosome 3p21-p14  
JOURNAL: Cytogenet. Cell Genet.  
VOLUME: 71  
ISSUE: 3  
PAGES: 296-298  
DATE: 1995  
DATABASE ACCESSION NUMBER: NM\_002218  
DATABASE ENTRY DATE: 2003-02-07  
RELEVANT RESIDUES: (1)..(930)  
US-10-444-575-6

Query Match 100.0%; Score 91; DB 30; Length 930;  
Best Local Similarity 100.0%; Pred. No. 4.6e-06;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSPROQETVVDGNLIIRY 18  
DB 224 KSPROQETVVDGNLIIRY 241

RESULT 16  
PCT-US01-14827-14552

Sequence 14552, Application PC/TUS0114827  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc  
TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
FILE REFERENCE: 21272-104  
CURRENT APPLICATION NUMBER: PCT/US01/14827  
CURRENT FILING DATE: 2001-05-16  
PRIOR APPLICATION NUMBER: 09/577,408  
PRIOR FILING DATE: 2000-05-18  
NUMBER OF SEQ ID NOS: 16102  
SOFTWARE: Custom  
SEQ ID NO 14552  
LENGTH: 971  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: DOMAIN  
LOCATION: (295)..(348)  
OTHER INFORMATION: INHIBITOR HEAVY CHAIN CHANNEL IN domain identified by  
OTHER INFORMATION: EMATRIX, accession number PD01101B, p-value=1.00e-40, raw score  
NAME/KEY: DOMAIN  
LOCATION: (313)..(496)  
OTHER INFORMATION: von Willebrand factor type A domain identified by Pfam,  
OTHER INFORMATION: accession name vwa, E-value=1.1e-08, Pfam score of 42.3  
PCT-US01-14827-14552

Query Match 100.0%; Score 91; DB 1; Length 971;  
Best Local Similarity 100.0%; Pred. No. 4.8e-06;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSPROQETVVDGNLIIRY 18  
DB 263 KSPROQETVVDGNLIIRY 280

RESULT 17  
PCT-US01-08656-10251

Sequence 10251, Application PC/TUS0108656  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc  
TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES  
FILE REFERENCE: 21272-066  
CURRENT APPLICATION NUMBER: PCT/US01/08656  
CURRENT FILING DATE: 2001-04-16

```

; PRIOR APPLICATION NUMBER: 09/522,929
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: 09/770,160
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 10994
; SOFTWARE: Custom
; SEQ ID NO: 10251
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-08656-10251

Query Match          94.5%; Score 86; DB 1; Length 123;
Best Local Similarity 94.4%; Pred. No. 2.5e-06;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 KSPEQOETVLDGNLIIRY 18
Db      53 KSPEQOETVLDGNLIIRY 70

RESULT 18
US-10-273-573-10251
; Sequence 10251, Application US/10273573
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-066
; CURRENT APPLICATION NUMBER: US/10/273,573
; CURRENT FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: 09/522,929
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: 09/770,160
; PRIOR FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 10994
; SOFTWARE: Custom
; SEQ ID NO: 10251
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-273-573-10251

Query Match          94.5%; Score 86; DB 28; Length 123;
Best Local Similarity 94.4%; Pred. No. 2.5e-06;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 KSPEQOETVLDGNLIIRY 18
Db      53 KSPEQOETVLDGNLIIRY 70

RESULT 19
US-09-993-295-2
; Sequence 2, Application US/09993295
; GENERAL INFORMATION:
; APPLICANT: Jackowski, George
; TITLE OF INVENTION: Macroclobulin Biopolymer Markers Indicative of Insulin Resistance
; FILE REFERENCE: 2132.099
; CURRENT APPLICATION NUMBER: US/09/993,295
; CURRENT FILING DATE: 2001-11-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 2
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-295-2

Query Match          88.5%; Score 80.5; DB 25; Length 19;
Best Local Similarity 94.7%; Pred. No. 2e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy      1 KSPE-QOETVLDGNLIIRY 18
```

```

Db      1 KSPEQOETVLDGNLIIRY 19

RESULT 20
PCT-US02-31642-356
; Sequence 356, Application PC/TUS0231642
; GENERAL INFORMATION:
; APPLICANT: OXFORD GLYSCOSCIENCES (UK) LTD.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES, POLYPEPTIDES AND USES THEREFOR, INCLUDING DIAGNOSIS AND TREATMENT OF ALZHEIMER'S DISEASE
; FILE REFERENCE: POA-002.25
; CURRENT APPLICATION NUMBER: PCT/US02/31642
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: 60/326708
; PRIOR FILING DATE: 2001-10-03
; NUMBER OF SEQ ID NOS: 491
; SOFTWARE: PatentIn Version 2.1
; SEQ ID NO: 356
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-31642-356

Query Match          86.8%; Score 79; DB 1; Length 16;
Best Local Similarity 100.0%; Pred. No. 2.9e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 SPEQOETVLDGNLIIR 17
Db      1 SPEQOETVLDGNLIIR 16

RESULT 21
US-09-826-290-44
; Sequence 44, Application US/09826290
; GENERAL INFORMATION:
; APPLICANT: Durham, L. Kathryn
; APPLICANT: Friedman, David L.
; APPLICANT: Herath, Herath Mudiyanselage Athula Chandrasiri
; APPLICANT: Kimmel, Lida H.
; APPLICANT: Parekh, Rajesh Bhikhu
; APPLICANT: Potter, David M.
; APPLICANT: Rohlf, Christian
; APPLICANT: Silber, B. Michael
; APPLICANT: Stieger, Thomas R.
; APPLICANT: Sunderland, P. Trey
; APPLICANT: Townsend, Robert Reid
; APPLICANT: White, Frost
; TITLE OF INVENTION: Nucleic Acid Molecules, Polypeptides and Uses Therefor, Including Diagnosis and Treatment of Alzheimer's Disease
; TITLE OF INVENTION: Alzheimer's Disease
; FILE REFERENCE: 2572-1-001 N2
; CURRENT APPLICATION NUMBER: US/09/826,290
; CURRENT FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: US 60/194,504
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 60/253,647
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 492
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 44
; LENGTH: 16
; TYPE: PRT
; ORGANISM: homo sapien
US-09-826-290-44

Query Match          86.8%; Score 79; DB 23; Length 16;
Best Local Similarity 100.0%; Pred. No. 2.9e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	2	SPEQQETVL	DGNL	IIR	17
Db	1	SPEQQETVL	DGNL	IIR	16

RESULT 22  
US-10-244-715A-356  
; Sequence 356, Application US/10244715A  
: GENERAL INFORMATION.

ORGANISM: Homo sapiens  
US-10-244-715A-356

```

1      RESULT 23
2      US-10-264-309-356
3      ; Sequence 356, Application US/10264309
4      ; GENERAL INFORMATION:
5      ; APPLICANT: DURHAM, L. KATHRYN
6      ; APPLICANT: FRIEDMAN, DAVID L.
7      ; APPLICANT: HERATH, HERATH
8      ; APPLICANT: KIMMEL, LIDA H.
9      ; APPLICANT: PAREKH, RAJESH B.
10     ; APPLICANT: POTTER, DAVID M.
11     ; APPLICANT: ROHLF, CHRISTIAN
12     ; APPLICANT: SILBER, B. MICHAEL
13     ; APPLICANT: SNYDER, PETER J.
14     ; APPLICANT: SOARES, HOLLY D.
15     ; APPLICANT: STIGER, THOMAS R.
16     ; APPLICANT: SUNDERLAND, P. TREY
17     ; APPLICANT: TOWNSEND, ROBERT R.
18     ; APPLICANT: WHITE, W. PROST
19     ; APPLICANT: WILLIAMS, STEPHEN A.
20     ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES, POLYPEPTIDES AND USES THEREFOR,
21     ; TITLE OF INVENTION: INCLUDING DIAGNOSIS AND TREATMENT OF ALZHEIMER'S DISEASE
22     ; FILE REFERENCE: PDA-002.01
23     ; CURRENT APPLICATION NUMBER: US/10/264,309
24     ; CURRENT FILING DATE: 2002-10-03
25     ; PRIORITY APPLICATION NUMBER: 60/326,708

```

```

; PRIOR FILING DATE: 2001-10-03
; NUMBER OF SEQ ID NOS: 491
; SOFTWARE: PatentIn Version 2.1.1
; SEQ ID NO 356
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-264-309-356

```

Query Match	86.8%;	Score 79;	DB 28;	Length 16;
Best Local Similarity	100.0%;	Pred. No. 2.9e-06;		
Matches 16;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

```

RESULT 24
US-10-624-429-281
; Sequence 281. Application US/10624429
; GENERAL INFORMATION:
; APPLICANT: Herath, Herath Mudiyanseelage Athula Chandrasiri
; APPLICANT: Parekh, Rajesh Bhikhu
; APPLICANT: Christian, Rohlf
; TITLE OF INVENTION: PROTEINS, GENES AND THEIR USE FOR DIAGNOSIS AND TREATMENT OF MULTIPLE SCLEROSIS
; TITLE OF INVENTION: SCLEROSIS
; FILE REFERENCE: 2543-1-002PCT CON
; CURRENT APPLICATION NUMBER: US/10/624,429
; CURRENT FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: PCT/GB02/00330
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: US 60/264404
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/331647
; PRIOR FILING DATE: 2001-11-20
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 281
; LENGTH: 16
; TYPE: prt
; ORGANISM: Homo sapiens
; US-10-624-429-281

```

```

Query Match      86.8%; Score 79; DB 31; Length 16;
Best Local Similarity 100.0%; Pred. No. 2,9e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qx 2 SPEQCEATVLDGNLILR 17
   |||||
Db 1 SPEQCEATVLDGNLILR 16

```

```

1 RESULT 25
2 US-10-444-575-4
3 ; Sequence 4, Application US/10444575
4 ; GENERAL INFORMATION:
5 ; APPLICANT: University of Connecticut Health Center
6 ; APPLICANT: Kuchel, George A
7 ; APPLICANT: Zhu, Qing
8 ; TITLE OF INVENTION: Compositions and Methods Relating to Detrusor Estrogen-Regulated
9 ; TITLE OF INVENTION: Protein (DERP)
10 ; FILE REFERENCE: UCT-0035
11 ; CURRENT APPLICATION NUMBER: US/10/444,575
12 ; CURRENT FILING DATE: 2003-05-22
13 ; PRIOR APPLICATION NUMBER: US 60/382,830
14 ; PRIOR FILING DATE: 2002-05-23
15 ; NUMBER OF SEQ ID NOS: 47
16 ; SOFTWARE: PatentIn version 3.2
17 ; SEQ ID NO 4
18 ; LENGTH: 932
19 ; TYPE: PRT
20 ; ORGANISM: Rattus norvegicus

```

```

; PUBLICATION INFORMATION:
; AUTHORS: Soury, E., Olivier, E., Daveau, M., Hiron, M., Claeysens, S.,
; AUTHORS: Ristler, J.L., and Sallier, J.P.
; TITLE: The H4p heavy chain of inter-alpha-inhibitor family largely
; TITLE: differs in the structure and synthesis of its proline-rich region
; TITLE: from rat to human
; JOURNAL: Biochem. Biophys. Res. Comm.
; VOLUME: 243
; ISSUE: 2
; PAGES: 522-530
; DATE: 1998-02-13
; DATABASE ACCESSION NUMBER: Y11283
; DATABASE ENTRY DATE: 1998-03-03
; RELEVANT RESIDUES: (1)..(932)
US-10-444-575-4

Query Match          62.6%; Score 57; DB 30; Length 932;
Best Local Similarity 55.6%; Pred. No. 4;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KSPQOETVLDGNLIIR 18
Db 223 KSONQDPTVLDGDDPTVRY 240

RESULT 26
US-10-444-575-1
; Sequence 1, Application US/10444575
; GENERAL INFORMATION:
; APPLICANT: University of Connecticut Health Center
; APPLICANT: Kuchel, George A
; APPLICANT: Zhu, Qing
; TITLE OF INVENTION: Compositions and Methods Relating to Detrusor Estrogen-Regulated
; TITLE OF INVENTION: Protein (DERP)
; FILE REFERENCE: UCT-0035
; CURRENT APPLICATION NUMBER: US/10/444,575
; CURRENT FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: US 60/382,830
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-444-575-1

Query Match          54.9%; Score 50; DB 30; Length 16;
Best Local Similarity 56.2%; Pred. No. 0.34;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SPEQOETVLDGNLIIR 17
Db 1 SONEQDPTVLDGDFIVR 16

RESULT 27
PCT-US02-38526-513
; Sequence 513, Application PC/TUS0238526
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Malabika
; APPLICANT: Tang, Y. Tom
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Zhao, Qing
; APPLICANT: Xu, Chongjun
; APPLICANT: Mulero, Julio J
; APPLICANT: Boyle, Bryan J.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL POLYPEPTIDES AND POLYNUCL
; FILE REFERENCE: HYS-B1/PCT
; CURRENT APPLICATION NUMBER: PCT/US02/38526
; CURRENT FILING DATE: 2002-12-03
; PRIOR APPLICATION NUMBER: US 09/488,725
```

```

; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: PCT/US00/34263
; PRIOR FILING DATE: 2000-12-26
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: PCT/US01/04098
; PRIOR FILING DATE: 2001-02-05
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 653
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 513
; LENGTH: 853
; TYPE: PRT
; ORGANISM: Bos taurus
PCT-US02-38526-513

Query Match          52.2%; Score 47.5; DB 1; Length 853;
Best Local Similarity 60.0%; Pred. No. 1.6e+02;
Matches 9; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

Qy 3 PEQOETVLDGNLIIR 17
Db 358 PEKQET-LDGHMVYR 371

RESULT 28
US-09-791-83086
; Sequence 83086, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBE
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 83086
; LENGTH: 853
; TYPE: PRT
; ORGANISM: Bos taurus
US-09-791-83086

Query Match          52.2%; Score 47.5; DB 22; Length 853;
Best Local Similarity 60.0%; Pred. No. 1.6e+02;
Matches 9; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

Qy 3 PEQOETVLDGNLIIR 17
Db 358 PEKQET-LDGHMVYR 371

RESULT 29
US-09-791-537-151450
; Sequence 151450, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Bionomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
```

```

; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 151450
; LENGTH: 853
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (556)..(556)
; OTHER INFORMATION: X is an unknown amino acid
US-09-791-537-151450

Query Match
Best Local Similarity 52.2%; Score 47.5; DB 22; Length 853;
Best Local Similarity 60.0%; Pred. No. 1.6e+02;
Matches 9; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

OY 3 PEOETVLIDGNLIIR 17
Db 358 PEOET-LDGHMVVR 371

RESULT 30
US-60-384-450-29
; Sequence 29, Application US/60384450
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Ghosh, Malabika
; APPLICANT: Mang, Jianrui
; APPLICANT: Hu, Tianhua
; TITLE OF INVENTION: Methods and Materials Relating to Neural Immunoglobulin Cell Adhe
; FILE REFERENCE: HYS-55
; CURRENT APPLICATION NUMBER: US/60/384,450
; CURRENT FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
; LENGTH: 853
; TYPE: PRT
; ORGANISM: Bos taurus
US-60-384-450-29

Query Match
Best Local Similarity 52.2%; Score 47.5; DB 32; Length 853;
Best Local Similarity 60.0%; Pred. No. 1.6e+02;
Matches 9; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

OY 3 PEOETVLIDGNLIIR 17
Db 358 PEOET-LDGHMVVR 371

RESULT 31
US-09-791-537-10439
; Sequence 10439, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Biomomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10439
; LENGTH: 858
; TYPE: PRT
```

```

; ORGANISM: Rattus norvegicus
US-09-791-537-10439

Query Match
Best Local Similarity 52.2%; Score 47.5; DB 22; Length 858;
Best Local Similarity 60.0%; Pred. No. 1.6e+02;
Matches 9; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

OY 3 PEOETVLIDGNLIIR 17
Db 360 PEOET-LDGHMVVR 373

RESULT 32
US-09-791-537-116884
; Sequence 116884, Application US/09791537
; GENERAL INFORMATION:
; APPLICANT: Biomomix, Inc.
; APPLICANT: Debe, Derek
; APPLICANT: Danzer, Joseph
; TITLE OF INVENTION: THREE DIMENSIONAL STRUCTURES OF PROTEIN FAMILIES AND FAMILY MEMBERS
; FILE REFERENCE: 261/210
; CURRENT APPLICATION NUMBER: US/09/791,537
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 153055
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 116884
; LENGTH: 858
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-791-537-116884

Query Match
Best Local Similarity 52.2%; Score 47.5; DB 22; Length 858;
Best Local Similarity 60.0%; Pred. No. 1.6e+02;
Matches 9; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

OY 3 PEOETVLIDGNLIIR 17
Db 360 PEOET-LDGHMVVR 373
```

```

RESULT 33
US-10-219-051B-9268
; Sequence 9268, Application US/10219051B
; GENERAL INFORMATION:
; APPLICANT: The General Hospital Corporation doing business as Massachusetts General
; APPLICANT: Hospital / Bayer AG
; TITLE OF INVENTION: Nucleotide sequences involved in pain
; FILE REFERENCE: Ica 35693 Foreign Countries
; CURRENT APPLICATION NUMBER: US/10/219,051B
; CURRENT FILING DATE: 2003-05-09
; PRIOR APPLICATION NUMBER: US 60/312,147
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: US 60/346,382
; PRIOR FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: US 60/333,347
; PRIOR FILING DATE: 2001-11-26
; NUMBER OF SEQ ID NOS: 14715
; SOFTWARE: Perl script
; SEQ ID NO 9268
; LENGTH: 858
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; DATABASE ACCESSION NUMBER: SWISS-Prot / P13596
; DATABASE ENTRY DATE: 2002-06-15
US-10-219-051B-9268

Query Match
Best Local Similarity 52.2%; Score 47.5; DB 28; Length 858;
Best Local Similarity 60.0%; Pred. No. 1.6e+02;
Matches 9; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

OY 3 PEOETVLIDGNLIIR 17
```

Db 360 PEQOET-LDGHMVVR 373

RESULT 34  
PCT-US02-38437-35

Sequence 35, Application PC/TUS0238437  
GENERAL INFORMATION:  
APPLICANT: INCYTE GENOMICS, INC.  
APPLICANT: BAUGHN, Marian R.  
APPLICANT: BECHA, Shanya D.  
APPLICANT: BHATIA, Umesh G.  
APPLICANT: BLAKE, Julie J.  
APPLICANT: BOROWSKY, Mark L.  
APPLICANT: BURRILL, John D.  
APPLICANT: DELEGEANE, Angelo M.  
APPLICANT: ELLIOTT, Vicki S.  
APPLICANT: GANDHI, Ameena R.  
APPLICANT: GIETZEN, Kimberly J.  
APPLICANT: GORVAD, Ann E.  
APPLICANT: GRIFFIN, Jennifer A.  
APPLICANT: HO, Anne  
APPLICANT: JIN, Pei  
APPLICANT: KABLE, Amy E.  
APPLICANT: LAL, Preeti G.  
APPLICANT: LEE, Ernestine A.  
APPLICANT: LEE, Sally  
APPLICANT: LEE, Soo Yuen  
APPLICANT: MARQUIS, Joseph P.  
APPLICANT: LEHR-MASON, Patricia M.  
APPLICANT: RAMKOMAR, Jayalakmi  
APPLICANT: RICHARDSON, Thomas W.  
APPLICANT: SPRAGUE, William W.  
APPLICANT: SMARNAKAR, Anita  
APPLICANT: TANG, Y. Tom  
APPLICANT: TRAN, Bao  
APPLICANT: TRAN, Uyen K.  
APPLICANT: CHAWLA, Narinder K.  
APPLICANT: WARREN, Bridget A.  
APPLICANT: XU, Yuming  
APPLICANT: YUE, Henry  
APPLICANT: ZHENG, Wenjin  
TITLE OF INVENTION: CELL ADHESION AND EXTRACELLULAR MATRIX PROTEINS  
FILE REFERENCE: PF-1296 PCT  
CURRENT APPLICATION NUMBER: PCT/US02/38437  
CURRENT FILING DATE: 2002-11-26  
PRIOR APPLICATION NUMBER: US 60/334,343  
PRIOR FILING DATE: 2001-11-30  
PRIOR APPLICATION NUMBER: US 60/340,278  
PRIOR FILING DATE: 2001-12-07  
PRIOR APPLICATION NUMBER: US 60/345,069  
PRIOR FILING DATE: 2002-01-04  
PRIOR APPLICATION NUMBER: US 60/351,352  
PRIOR FILING DATE: 2002-01-25  
PRIOR APPLICATION NUMBER: US 60/357,168  
PRIOR FILING DATE: 2002-02-14  
PRIOR APPLICATION NUMBER: US 60/369,128  
PRIOR FILING DATE: 2002-03-29  
PRIOR APPLICATION NUMBER: US 60/370,802  
PRIOR FILING DATE: 2002-04-05  
NUMBER OF SEQ ID NOS: 74  
SOFTWARE: PERL Program  
SEQ ID NO 35  
LENGTH: 859  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No: 7510063CD1  
PCT-US02-38437-35

Query Match 52.2%; Score 47.5; DB 1; Length 859;  
Best Local Similarity 60.0%; Pred. No. 1.6e+02;

Matches 9; Conservative 5; Mismatches 0; Indels 1; Gaps 1;  
Qy 3 PEQOETVLDGNLIIR 17  
Db 359 PEQOET-LDGHMVVR 372

RESULT 35  
US-10-156-761-14851  
Sequence 14851, Application US/10156761  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 14851  
LENGTH: 601  
TYPE: PRT  
ORGANISM: Streptomyces avermectilis  
US-10-156-761-14851

Query Match 50.5%; Score 46; DB 27; Length 601;  
Best Local Similarity 53.3%; Pred. No. 1.9e+02;  
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 3 PEQOETVLDGNLIIR 17  
Db 57 PERQTAVDGHLIR 71

RESULT 36  
US-10-424-599-177890  
Sequence 177890, Application US/10424599  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: KOVALLIC David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 177890  
LENGTH: 114  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_131650C.1.pep  
US-10-424-599-177890

Query Match 49.5%; Score 45; DB 30; Length 114;  
Best Local Similarity 53.8%; Pred. No. 32;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 3 PEQOETVLDGNLI 15  
Db 18 PDQOPLIFDGNL 30

```
RESULT 37
US-09-708-427-6340
; Sequence 6340, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6340
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..428
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc feature
; LOCATION: 1..428
; OTHER INFORMATION: Ceres Seq. ID 1815739
US-09-708-427-6340

Query Match          49.5%; Score 45; DB 21; Length 428;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 KSPEQETVLDGNLIIRY 18
Db 136 KDEPKQDKVLEGHPLRY 153

RESULT 38
US-09-708-427-6339
; Sequence 6339, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6339
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..429
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc feature
; LOCATION: 1..429
; OTHER INFORMATION: Ceres Seq. ID 1815738
US-09-708-427-6339

Query Match          49.5%; Score 45; DB 21; Length 429;
Best Local Similarity 50.0%; Pred. No. 1.8e+02;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 KSPEQETVLDGNLIIRY 18
Db 137 KDEPKQDKVLEGHPLRY 154

RESULT 39
US-09-708-427-6338
; Sequence 6338, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
```

```
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES 1
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6338
; LENGTH: 452
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1..452
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc feature
; LOCATION: 1..452
; OTHER INFORMATION: Ceres Seq. ID 1815737
US-09-708-427-6338

Query Match          49.5%; Score 45; DB 21; Length 452;
Best Local Similarity 50.0%; Pred. No. 1.9e+02;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 KSPEQETVLDGNLIIRY 18
Db 160 KDEPKQDKVLEGHPLRY 177

RESULT 40
US-10-437-963-129305
; Sequence 129305, Application US/10437963
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovacic, David K.
; APPLICANT: Zhou, Yinhua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21 (53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 129305
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(81)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_31574C.1.pep
US-10-437-963-129305

Query Match          48.4%; Score 44; DB 30; Length 81;
Best Local Similarity 69.2%; Pred. No. 31;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KSPEQETVLDGN 13
Db 53 KSPLOESTVLGDN 65

Search completed: October 10, 2003, 17:33:00
Job time : 408 secs
```

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 10, 2003, 17:23:09 ; Search time 25 Seconds

(without alignments)  
21.369 Million cell updates/sec

Title: US-09-991-795-1

Sequence: 1 KSPBQETVLDGNLIRY 18

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 143239 seqs, 29679805 residues

Total number of hits satisfying chosen parameters: 143239

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

Pending Patents\_AA New:\*  
1: /cgn2\_6/ptodata/2/paa/PCT\_US06\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	91	100.0	18	5	US-09-991-795-1
2	91	100.0	748	1	PCT-US03-23249-3
3	91	100.0	845	1	PCT-US03-00252A-8
4	91	100.0	845	1	PCT-US03-28227-3991
5	91	100.0	873	1	PCT-US03-28227-3990
6	91	100.0	882	1	PCT-US03-28227-3989
7	91	100.0	882	1	PCT-US03-00252A-10
8	91	100.0	885	1	PCT-US03-28227-3988
9	91	100.0	891	1	PCT-US03-23249-4
10	91	100.0	900	1	PCT-US03-00252A-12
11	91	100.0	930	1	PCT-US03-00252A-6
12	43	47.3	640	6	US-10-425-114A-43992
13	42	46.2	667	6	US-10-425-114A-46636
14	42	46.2	698	6	US-10-425-114A-68826
15	40.5	44.5	557	7	US-60-485-450-1055
16	40.5	44.5	557	7	US-60-485-450-1055
17	40	44.0	366	1	PCT-US03-20001-38
18	40	44.0	366	1	PCT-US03-20001-40
19	40	44.0	384	1	PCT-US03-20001-42
20	40	44.0	384	1	PCT-US03-20001-44
21	40	44.0	623	6	US-10-469-204-167
22	39	42.9	1034	1	PCT-US03-29089-28
23	39	42.9	1132	7	US-60-485-450-1641
24	38	41.8	109	1	PCT-US03-28508-59
25	38	41.8	209	6	US-10-425-114A-64443
26	38	41.8	294	5	US-09-897-516A-5308

27	38	41.8	353	7	US-60-485-114-2249	Sequence 2249, Ap
28	38	41.8	356	6	US-10-644-807-430	Sequence 430, Ap
29	38	41.8	359	1	PCT-US03-26780-1271	Sequence 1271, Ap
30	38	41.8	365	6	US-10-644-807-344	Sequence 344, Ap
31	38	41.8	367	6	US-10-425-114A-62212	Sequence 62212, A
32	38	41.8	377	7	US-60-485-450-1641	Sequence 1641, Ap
33	38	41.8	384	1	PCT-US03-20001-32	Sequence 32, Ap
34	38	41.8	385	1	PCT-US03-20001-51	Sequence 51, Ap
35	38	41.8	385	1	PCT-US03-20001-28	Sequence 28, Ap
36	38	41.8	392	1	PCT-US03-20001-30	Sequence 30, Ap
37	38	41.8	393	7	US-60-485-450-1640	Sequence 1640, Ap
38	38	41.8	423	7	US-60-485-450-1638	Sequence 1638, Ap
39	38	41.8	439	7	US-60-485-450-1639	Sequence 1639, Ap
40	38	41.8	452	7	US-60-485-114-1338	Sequence 1338, Ap
41	38	41.8	456	6	US-10-425-114A-68963	Sequence 68963, A
42	38	41.8	466	6	US-10-644-807-250	Sequence 250, Ap
43	38	41.8	516	7	US-60-485-114-2250	Sequence 2250, Ap
44	38	41.8	554	6	US-10-469-204-161	Sequence 161, Ap
45	38	41.8	881	6	US-10-425-114A-65404	Sequence 65404, A

## ALIGNMENTS

```
RESULT 1
US-09-991-795-1
Sequence 1, Application US/09991795
GENERAL INFORMATION:
APPLICANT: Jackowski, George
TITLE OF INVENTION: Inter Alpha Trypsin Inhibitor Biopolymer Markers Indicative of In-
FILE REFERENCE: 2132.105
CURRENT APPLICATION NUMBER: US/09/991.795
CURRENT FILING DATE: 2001-11-23
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 18
TYPE: PRT
ORGANISM: Homo sapiens
US-09-991-795-1
Query Match      100.0%; Score 91; DB 5; Length 18;
Best Local Similarity 100.0%; Pred. No. 3e-05;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy      1      KSPBQETVLDGNLIRY 18
Db      1      KSPBQETVLDGNLIRY 18
RESULT 2
PCT-US03-23249-3
Sequence 3, Application PC/TUS0323249
GENERAL INFORMATION:
APPLICANT: INCYTE CORPORATION
APPLICANT: ELLIOTT, Vicki S.; SWARNKAR, Anita;
APPLICANT: GRIFFIN, Jennifer A.; LEE, Ernestine A.;
APPLICANT: SPRAGUE, William W.; HAPPLA, April J.A.;
APPLICANT: LEE, Soo Yeun; KABER, Amy E.;
APPLICANT: TSON, Craig H.; KEARE, Reena;
APPLICANT: CHAWLA, Narinder K.; MARQUIS, Joseph P.;
APPLICANT: JIANG Xin; JACKSON, Alan A.;
APPLICANT: BECHA, Shanya D.; EMBERLING, Brooke M.;
APPLICANT: JIN, Pei; WILSON, Amy D.;
APPLICANT: RICHARDSON, Thomas W.; YANG, Junling;
APPLICANT: BAUGHN, Mariah R.; GANDHI, Ameena R.;
APPLICANT: NGUYEN, Daniel B.; RAMKOMAR, Jayalaxmi;
APPLICANT: KALLICK, Deborah A.; KEARNEY, Liam;
APPLICANT: LU, Dying Aina M.; GIETZEN, Kimberly J.;
APPLICANT: TRIBOUTEY, Catherine M.; LAL, Preeti G.;
APPLICANT: BLAKE, Julie J.; LU, Yan;
```

```
APPLICANT: ARVIZU, Chandra S.
TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
FILE REFERENCE: PF-1478 PCT
CURRENT APPLICATION NUMBER: PCT/US03/23249
CURRENT FILING DATE: 2003-07-23
PRIOR APPLICATION NUMBER: US 60/398,143
PRIOR FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 60/402,458
PRIOR FILING DATE: 2002-08-09
PRIOR APPLICATION NUMBER: US 60/403,289
PRIOR FILING DATE: 2002-08-12
PRIOR APPLICATION NUMBER: US 60/406,472
PRIOR FILING DATE: 2002-08-27
PRIOR APPLICATION NUMBER: US 60/409,354
PRIOR FILING DATE: 2002-09-06
NUMBER OF SEQ ID NOS: 124
SOFTWARE: PERL Program
SEQ ID NO 3
LENGTH: 748
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No: 7512557CD1
PCT-US03-23249-3
```

```
Query Match          100.0%; Score 91; DB 1; Length 748;
Best Local Similarity 100.0%; Pred. No. 0.00029;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 KSPSEQETVLDGNLIIRY 18
        |||||||
Db      224 KSPSEQETVLDGNLIIRY 241
```

```
RESULT 3
PCT-US03-00252A-8
Sequence 8, Application PC/TUS0300252A
GENERAL INFORMATION:
APPLICANT: Curagen Corporation
TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
FILE REFERENCE: 21402-533A-061
CURRENT APPLICATION NUMBER: PCT/US03/00252A
CURRENT FILING DATE: 2003-01-06
PRIOR APPLICATION NUMBER: 10/336,603
PRIOR FILING DATE: 2003-01-03
NUMBER OF SEQ ID NOS: 169
SOFTWARE: Curaseqdist version 0.1
SEQ ID NO 8
LENGTH: 843
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US03-00252A-8
```

```
Query Match          100.0%; Score 91; DB 1; Length 843;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 KSPSEQETVLDGNLIIRY 18
        |||||||
Db      224 KSPSEQETVLDGNLIIRY 241
```

```
RESULT 4
PCT-US03-28227-3991
Sequence 3991, Application PC/TUS0328227
GENERAL INFORMATION:
APPLICANT: INCYTE CORPORATION; SCHMIDT, Jeanette P.;
APPLICANT: WRIGHT, Rachel J.; BRUNS, Christopher M.;
APPLICANT: MARJANOVIC, Mirjana M.; SHEN, Fan;
APPLICANT: HARTSHORNE, ToINETTE A.; SUCHOROLSKI, Martin;
APPLICANT: ALTUS, Christina M.; PITTS, Steven J.;
APPLICANT: ELDER, Linda V.; MOONEY, Elizabeth M.;
```

```
APPLICANT: DELEGANE, Angelo M.; PANESAR, Iqbal S.;
APPLICANT: BANVILLE, Steven C.; REDDY, Thirupathi P.;
APPLICANT: STEVENS, Kristian A.; BLANCHARD, John L.;
APPLICANT: PANZER, Scott R.; WANG, Xinhao;
APPLICANT: AU, Alan P.; GERSTIN, Edward H., Jr.;
APPLICANT: PERALTA, Careyna H.; ANDERSON, Scott E.;
APPLICANT: RIOUX, Pierre; SHEN, Edward J.;
APPLICANT: WU, Mingham C.; STUVE, Laura L.;
APPLICANT: LAGACE, Robert E.; SPIRO, Peter A.;
APPLICANT: STEWART, Elizabeth A.; WINGROVE, James A.;
APPLICANT: VITT, Ursula A.; KIRTON, Edward;
APPLICANT: XU, Yuming; KWONG, Mary;
APPLICANT: POLICKY, Jennifer L.; HURWITZ, Bonnie L.;
APPLICANT: MA, Yan; JACKSON, Jennifer L.;
APPLICANT: GIETZEN, Darryl; PATURY, Srikanth;
APPLICANT: SHI, Xiaobing; SUAREZ, Charilyn J.
TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: PN-0100 PCT
CURRENT APPLICATION NUMBER: PCT/US03/28227
CURRENT FILING DATE: 2003-09-12
PRIOR APPLICATION NUMBER: US 60/410,260
PRIOR FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 60/410,259
PRIOR FILING DATE: 2002-09-12
NUMBER OF SEQ ID NOS: 5444
SOFTWARE: PERL Program
SEQ ID NO 3991
LENGTH: 845
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No: 944084.PT1904P
PCT-US03-28227-3991
```

```
Query Match          100.0%; Score 91; DB 1; Length 845;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 KSPSEQETVLDGNLIIRY 18
        |||||||
Db      224 KSPSEQETVLDGNLIIRY 241
```

```
RESULT 5
PCT-US03-28227-3990
Sequence 3990, Application PC/TUS0328227
GENERAL INFORMATION:
APPLICANT: INCYTE CORPORATION; SCHMIDT, Jeanette P.;
APPLICANT: WRIGHT, Rachel J.; BRUNS, Christopher M.;
APPLICANT: MARJANOVIC, Mirjana M.; SHEN, Fan;
APPLICANT: HARTSHORNE, ToINETTE A.; SUCHOROLSKI, Martin;
APPLICANT: ALTUS, Christina M.; PITTS, Steven J.;
APPLICANT: ELDER, Linda V.; MOONEY, Elizabeth M.;
APPLICANT: DELEGANE, Angelo M.; PANESAR, Iqbal S.;
APPLICANT: BANVILLE, Steven C.; REDDY, Thirupathi P.;
APPLICANT: STEVENS, Kristian A.; BLANCHARD, John L.;
APPLICANT: PANZER, Scott R.; WANG, Xinhao;
APPLICANT: AU, Alan P.; GERSTIN, Edward H., Jr.;
APPLICANT: PERALTA, Careyna H.; ANDERSON, Scott E.;
APPLICANT: WU, Mingham C.; STUVE, Laura L.;
APPLICANT: LAGACE, Robert E.; SPIRO, Peter A.;
APPLICANT: STEWART, Elizabeth A.; WINGROVE, James A.;
APPLICANT: VITT, Ursula A.; KIRTON, Edward;
APPLICANT: XU, Yuming; KWONG, Mary;
APPLICANT: POLICKY, Jennifer L.; HURWITZ, Bonnie L.;
APPLICANT: MA, Yan; JACKSON, Jennifer L.;
APPLICANT: GIETZEN, Darryl; PATURY, Srikanth;
APPLICANT: SHI, Xiaobing; SUAREZ, Charilyn J.
TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: PN-0100 PCT
CURRENT APPLICATION NUMBER: PCT/US03/28227
```

```

CURRENT FILING DATE: 2003-09-12
PRIOR APPLICATION NUMBER: US 60/410,260
PRIOR FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 60/410,259
PRIOR FILING DATE: 2002-09-12
NUMBER OF SEQ ID NOS: 5444
SOFTWARE: PERL Program
SEQ ID NO 3990
LENGTH: 873
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 944084.PT1903p
PCT-US03-28227-3990

Query Match      100.0%; Score 91; DB 1; Length 873;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 KSPQEQETVLDGNLIIRY 18
        |||||
Db      224 KSPQEQETVLDGNLIIRY 241

RESULT 6
PCT-US03-28227-3989
Sequence 3989, Application PC/TUS0328227
GENERAL INFORMATION:
APPLICANT: INCYTE CORPORATION; SCHMIDT, Jeanette P.;
APPLICANT: WRIGHT, Rachel J.; BRUNS, Christopher M.;
APPLICANT: MARJANOVIC, Mirjana M.; SHEN, Fan;
APPLICANT: HARTSHORNE, Tolnette A.; SUCHOROLSKI, Martin;
APPLICANT: ALTUS, Christina M.; PITTS, Steven J.;
APPLICANT: ELDER, Linda V.; MOONEY, Elizabeth M.;
APPLICANT: DELEGANE, Angelo M.; PANESAR, Iqbal S.;
APPLICANT: BANVILLE, Steven C.; REDDY, Thirupathi P.;
APPLICANT: STEVENS, Kristian A.; BLANCHARD, John L.;
APPLICANT: PANZER, Scott R.; WANG, Xinhao;
APPLICANT: AU, Alan P.; GERSTIN, Edward H., Jr.;
APPLICANT: PERALTA, Careyana H.; ANDERSON, Scott E.;
APPLICANT: RIOUX, Pierre; SHEN, Edward J.;
APPLICANT: WU, Mingham C.; STUVE, Laura L.;
APPLICANT: LAGACE, Robert E.; SPIRO, Peter A.;
APPLICANT: STEWART, Elizabeth A.; WINGROVE, James A.;
APPLICANT: VITT, Ursula A.; KIRTON, Edward;
APPLICANT: XU, Yuming; KWONG, Mary;
APPLICANT: POLICKY, Jennifer L.; HURWITZ, Bonnie L.;
APPLICANT: MA, Yan; JACKSON, Jennifer L.;
APPLICANT: GIETZEN, Darryl; PATURY, Srikanth;
APPLICANT: SHI, Xiaobing; SUAREZ, Charlyan J.
TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: PN-0100 PCT
CURRENT APPLICATION NUMBER: PCT/US03/28227
CURRENT FILING DATE: 2003-09-12
PRIOR APPLICATION NUMBER: US 60/410,260
PRIOR FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 60/410,259
PRIOR FILING DATE: 2002-09-12
NUMBER OF SEQ ID NOS: 5444
SOFTWARE: PERL Program
SEQ ID NO 3989
LENGTH: 882
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 944084.PT1792p
PCT-US03-28227-3989
```

```

Query Match      100.0%; Score 91; DB 1; Length 882;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

Qy      1 KSPQEQETVLDGNLIIRY 18
        |||||
Db      224 KSPQEQETVLDGNLIIRY 241

RESULT 7
PCT-US03-00252A-10
Sequence 10, Application PC/TUS0300252A
GENERAL INFORMATION:
APPLICANT: Curagen Corporation
TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
FILE REFERENCE: 21402-533A-061
CURRENT APPLICATION NUMBER: PCT/US03/00252A
CURRENT FILING DATE: 2003-01-06
PRIOR APPLICATION NUMBER: 10/336,603
PRIOR FILING DATE: 2003-01-03
NUMBER OF SEQ ID NOS: 169
SOFTWARE: Curaseqlist version 0.1
SEQ ID NO 10
LENGTH: 882
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US03-00252A-10

Query Match      100.0%; Score 91; DB 1; Length 882;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

Qy      1 KSPQEQETVLDGNLIIRY 18
        |||||
Db      224 KSPQEQETVLDGNLIIRY 241
```

```

RESULT 8
PCT-US03-28227-3988
Sequence 3988, Application PC/TUS0328227
GENERAL INFORMATION:
APPLICANT: INCYTE CORPORATION; SCHMIDT, Jeanette P.;
APPLICANT: WRIGHT, Rachel J.; BRUNS, Christopher M.;
APPLICANT: MARJANOVIC, Mirjana M.; SHEN, Fan;
APPLICANT: HARTSHORNE, Tolnette A.; SUCHOROLSKI, Martin;
APPLICANT: ALTUS, Christina M.; PITTS, Steven J.;
APPLICANT: ELDER, Linda V.; MOONEY, Elizabeth M.;
APPLICANT: DELEGANE, Angelo M.; PANESAR, Iqbal S.;
APPLICANT: BANVILLE, Steven C.; REDDY, Thirupathi P.;
APPLICANT: STEVENS, Kristian A.; BLANCHARD, John L.;
APPLICANT: PANZER, Scott R.; WANG, Xinhao;
APPLICANT: AU, Alan P.; GERSTIN, Edward H., Jr.;
APPLICANT: PERALTA, Careyana H.; ANDERSON, Scott E.;
APPLICANT: RIOUX, Pierre; SHEN, Edward J.;
APPLICANT: WU, Mingham C.; STUVE, Laura L.;
APPLICANT: LAGACE, Robert E.; SPIRO, Peter A.;
APPLICANT: STEWART, Elizabeth A.; WINGROVE, James A.;
APPLICANT: VITT, Ursula A.; KIRTON, Edward;
APPLICANT: XU, Yuming; KWONG, Mary;
APPLICANT: POLICKY, Jennifer L.; HURWITZ, Bonnie L.;
APPLICANT: MA, Yan; JACKSON, Jennifer L.;
APPLICANT: GIETZEN, Darryl; PATURY, Srikanth;
APPLICANT: SHI, Xiaobing; SUAREZ, Charlyan J.
TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: PN-0100 PCT
CURRENT APPLICATION NUMBER: PCT/US03/28227
CURRENT FILING DATE: 2003-09-12
PRIOR APPLICATION NUMBER: US 60/410,260
PRIOR FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 60/410,259
PRIOR FILING DATE: 2002-09-12
NUMBER OF SEQ ID NOS: 5444
SOFTWARE: PERL Program
SEQ ID NO 3988
LENGTH: 885
TYPE: PRT
```

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 944084.PT1788p
PCT-US03-28227-3988

Query Match
Best Local Similarity 100.0%; Score 91; DB 1; Length 885;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSPEQETVLDGNLIIRY 18
   |||||
Db 224 KSPEQETVLDGNLIIRY 241

RESULT 9
PCT-US03-23249-4
; Sequence 4, Application PC/TUS0323249
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION
; APPLICANT: ELIOTT, Vicki S.; SWANNAKAR, Anita;
; APPLICANT: GRIFFIN, Jennifer A.; LEE, Ernestine A.;
; APPLICANT: SPRAGUE, William W.; HAFALIA, April J.A.;
; APPLICANT: LEE, Soo Yeun; KABLE, Amy E.;
; APPLICANT: ISON, Craig H.; KHARE, Reena;
; APPLICANT: CHAMLA, Narinder K.; MARQUIS, Joseph P.;
; APPLICANT: JIANG, Xin; JACKSON, Alan A.;
; APPLICANT: BECHA, Shanya D.; EMERLING, Brooke M.;
; APPLICANT: JIN, Pei; WILSON, Amy D.;
; APPLICANT: RICHARDSON, Thomas W.; YANG, Junming;
; APPLICANT: BAUGHN, Mariah R.; GANDHI, Ameena R.;
; APPLICANT: NGUYEN, Daniel B.; RAKUTOMAR, Jayalaxmi;
; APPLICANT: KALITCK, Deborah A.; KEARNEY, Liam;
; APPLICANT: LU, Dyanq, Aina M.; GIEREN, Kimberly J.;
; APPLICANT: TRIBOULEY, Catherine M.; LAL, Preeti G.;
; APPLICANT: BLAKE, Julie J.; LU, Yan;
; APPLICANT: ARVIZU, Chandra S.
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PR-1478 PCT
; CURRENT APPLICATION NUMBER: PCT/US03/23249
; CURRENT FILING DATE: 2003-07-23
; PRIOR APPLICATION NUMBER: US 60/398,143
; PRIOR FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/402,458
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/403,289
; PRIOR FILING DATE: 2002-08-12
; PRIOR APPLICATION NUMBER: US 60/406,472
; PRIOR FILING DATE: 2002-08-27
; PRIOR APPLICATION NUMBER: US 60/409,354
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 891
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 7512559CD1
PCT-US03-23249-4

Query Match
Best Local Similarity 100.0%; Score 91; DB 1; Length 891;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSPEQETVLDGNLIIRY 18
   |||||
Db 224 KSPEQETVLDGNLIIRY 241

RESULT 10
PCT-US03-00252A-12

; Sequence 12, Application PC/TUS0300252A
; GENERAL INFORMATION:
; APPLICANT: Curagen Corporation
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-533A-061
; CURRENT APPLICATION NUMBER: PCT/US03/00252A
; CURRENT FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: 10/336,603
; PRIOR FILING DATE: 2003-01-03
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: Cursesqlist version 0.1
; SEQ ID NO 12
; LENGTH: 900
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US03-00252A-12

Query Match
Best Local Similarity 100.0%; Score 91; DB 1; Length 900;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSPEQETVLDGNLIIRY 18
   |||||
Db 224 KSPEQETVLDGNLIIRY 241

RESULT 11
PCT-US03-00252A-6
; Sequence 6, Application PC/TUS0300252A
; GENERAL INFORMATION:
; APPLICANT: Curagen Corporation
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-533A-061
; CURRENT APPLICATION NUMBER: PCT/US03/00252A
; CURRENT FILING DATE: 2003-01-06
; PRIOR APPLICATION NUMBER: 10/336,603
; PRIOR FILING DATE: 2003-01-03
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: Cursesqlist version 0.1
; SEQ ID NO 6
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US03-00252A-6

Query Match
Best Local Similarity 100.0%; Score 91; DB 1; Length 930;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSPEQETVLDGNLIIRY 18
   |||||
Db 224 KSPEQETVLDGNLIIRY 241

RESULT 12
US-10-425-114A-43992
; Sequence 43992, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114A
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 43992
; LENGTH: 640
; TYPE: PRT
```

```

; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700685085_FLI.dep
US-10-425-114A-43992

Query Match          47.3%; Score 43; DB 6; Length 640;
Best Local Similarity 69.2%; Pred. No. 1.2e+02;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY      2 SPCQETVLNDGL 14
Db      554 SPCQRTMLGENL 566

RESULT 13
US-10-425-114A-46636
; Sequence 46636, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT FILING DATE: 2003-04-28
; CURRENT APPLICATION NUMBER: US/10/425, 114A
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 46636
; LENGTH: 667
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700093518_FLI.dep
US-10-425-114A-46636

Query Match          46.2%; Score 42; DB 6; Length 667;
Best Local Similarity 61.5%; Pred. No. 1.7e+02;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY      2 SPCQETVLNDGL 14
Db      580 NPEQRTILGENL 592

RESULT 14
US-10-425-114A-68826
; Sequence 68826, Application US/10425114A
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 68826
; LENGTH: 698
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73200B10_FLI.dep
US-10-425-114A-68826

Query Match          46.2%; Score 42; DB 6; Length 698;
Best Local Similarity 61.5%; Pred. No. 1.7e+02;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY      2 SPCQETVLNDGL 14
Db      580 NPEQRTILGENL 592

MATCHES 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY      2 SPCQETVLNDGL 14
Db      611 NPEQRTILGENL 623

RESULT 15
US-60-487-610-1675
; Sequence 1675, Application US/60487610
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: HUANG, Hongjin
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: LIVER FIBROSIS IN HEPATITIS C VIRUS-INFECTED SUBJECTS,
; TITLE OF INVENTION: METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01469
; CURRENT FILING DATE: 2003-07-17
; NUMBER OF SEQ ID NOS: 97101
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1675
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-487-610-1675

Query Match          44.5%; Score 40.5; DB 7; Length 557;
Best Local Similarity 42.1%; Pred. No. 2.2e+02;
Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

OY      1 KSPQKQET-VLDGLIIRY 18
Db      31 KSPQKVEVDIIDNFIILRW 49

RESULT 16
US-60-485-450-1055
; Sequence 1055, Application US/60485450
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: CHANG, Sheng-Yung
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RESPONSE TO INTERFERON TREATMENT IN HEPATITIS C
; TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO01470
; CURRENT FILING DATE: 2003-07-09
; NUMBER OF SEQ ID NOS: 47859
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1055
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-485-450-1055

Query Match          44.5%; Score 40.5; DB 7; Length 557;
Best Local Similarity 42.1%; Pred. No. 2.2e+02;
Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

OY      1 KSPQKQET-VLDGLIIRY 18
Db      31 KSPQKVEVDIIDNFIILRW 49

RESULT 17
PCT-US03-20001-38
; Sequence 38, Application PC/TUS0320001
; GENERAL INFORMATION:
; APPLICANT: Boyes, Douglas
; APPLICANT: Davis, Keith
; APPLICANT: Jones, Alan
```

```
APPLICANT: Ullah, Hemayet
APPLICANT: Chen, Jin-Gui
APPLICANT: Mulpuri, Rao
APPLICANT: Chatterjee, Anil
APPLICANT: Ward, Mary
TITLE OF INVENTION: METHODS FOR IMPROVING PLANT AGRONOMICAL TRAITS BY ALTERING THE
TITLE OF INVENTION: EXPRESSION OR ACTIVITY OF PLANT G-PROTEIN ALPHA AND BETA SUBUNIT
FILE REFERENCE: 2155US
CURRENT APPLICATION NUMBER: PCT/US03/20001
CURRENT FILING DATE: 2003-06-24
PRIOR APPLICATION NUMBER: 60/392,730
PRIOR FILING DATE: 2002-06-28
PRIOR APPLICATION NUMBER: 60/445,208
PRIOR FILING DATE: 2003-02-05
NUMBER OF SEQ ID NOS: 78
SOFTWARE: PatentIn version 3.2
SEQ ID NO 38
LENGTH: 366
TYPE: PRT
ORGANISM: Nicotiana tomentosiformis
PCT-US03-20001-38
```

```
Query Match          44.0% Score 40; DB 1; Length 366;
Best Local Similarity 61.5% Pred. No. 2e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      1 KSPSEQEVTLDGN 13
      |||:|:|
Db      134 KDPALQETILRGN 146
```

```
RESULT 18
PCT-US03-20001-40
; Sequence 40, Application PC/TUS0320001
; GENERAL INFORMATION:
```

```
APPLICANT: Boyes, Douglas
APPLICANT: Davis, Keith
APPLICANT: Jones, Alan
APPLICANT: Ullah, Hemayet
APPLICANT: Chen, Jin-Gui
APPLICANT: Mulpuri, Rao
APPLICANT: Chatterjee, Anil
APPLICANT: Ward, Mary
TITLE OF INVENTION: METHODS FOR IMPROVING PLANT AGRONOMICAL TRAITS BY ALTERING THE
TITLE OF INVENTION: EXPRESSION OR ACTIVITY OF PLANT G-PROTEIN ALPHA AND BETA SUBUNIT
FILE REFERENCE: 2155US
CURRENT APPLICATION NUMBER: PCT/US03/20001
CURRENT FILING DATE: 2003-06-24
PRIOR APPLICATION NUMBER: 60/392,730
PRIOR FILING DATE: 2002-06-28
PRIOR APPLICATION NUMBER: 60/445,208
PRIOR FILING DATE: 2003-02-05
NUMBER OF SEQ ID NOS: 78
SOFTWARE: PatentIn version 3.2
SEQ ID NO 40
LENGTH: 366
TYPE: PRT
ORGANISM: Nicotiana tabacum
PCT-US03-20001-40
```

```
Query Match          44.0% Score 40; DB 1; Length 366;
Best Local Similarity 61.5% Pred. No. 2e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      1 KSPSEQEVTLDGN 13
      |||:|:|
Db      134 KDPALQETILRGN 146
```

```
RESULT 19
PCT-US03-20001-42
; Sequence 42, Application PC/TUS0320001
; GENERAL INFORMATION:
```

```
APPLICANT: Boyes, Douglas
APPLICANT: Davis, Keith
APPLICANT: Jones, Alan
APPLICANT: Ullah, Hemayet
APPLICANT: Chen, Jin-Gui
APPLICANT: Mulpuri, Rao
APPLICANT: Chatterjee, Anil
APPLICANT: Ward, Mary
TITLE OF INVENTION: METHODS FOR IMPROVING PLANT AGRONOMICAL TRAITS BY ALTERING THE
TITLE OF INVENTION: EXPRESSION OR ACTIVITY OF PLANT G-PROTEIN ALPHA AND BETA SUBUNIT
FILE REFERENCE: 2155US
CURRENT APPLICATION NUMBER: PCT/US03/20001
CURRENT FILING DATE: 2003-06-24
PRIOR APPLICATION NUMBER: 60/392,730
PRIOR FILING DATE: 2002-06-28
PRIOR APPLICATION NUMBER: 60/445,208
PRIOR FILING DATE: 2003-02-05
NUMBER OF SEQ ID NOS: 78
SOFTWARE: PatentIn version 3.2
SEQ ID NO 42
LENGTH: 384
TYPE: PRT
ORGANISM: Nicotiana tabacum
PCT-US03-20001-42
```

```
Query Match          44.0% Score 40; DB 1; Length 384;
Best Local Similarity 61.5% Pred. No. 2.1e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      1 KSPSEQEVTLDGN 13
      |||:|:|
Db      145 KDPALQETILRGN 157
```

```
RESULT 20
PCT-US03-20001-44
; Sequence 44, Application PC/TUS0320001
; GENERAL INFORMATION:
```

```
APPLICANT: Boyes, Douglas
APPLICANT: Davis, Keith
APPLICANT: Jones, Alan
APPLICANT: Ullah, Hemayet
APPLICANT: Chen, Jin-Gui
APPLICANT: Mulpuri, Rao
APPLICANT: Chatterjee, Anil
APPLICANT: Ward, Mary
TITLE OF INVENTION: METHODS FOR IMPROVING PLANT AGRONOMICAL TRAITS BY ALTERING THE
TITLE OF INVENTION: EXPRESSION OR ACTIVITY OF PLANT G-PROTEIN ALPHA AND BETA SUBUNIT
FILE REFERENCE: 2155US
CURRENT APPLICATION NUMBER: PCT/US03/20001
CURRENT FILING DATE: 2003-06-24
PRIOR APPLICATION NUMBER: 60/392,730
PRIOR FILING DATE: 2002-06-28
PRIOR APPLICATION NUMBER: 60/445,208
PRIOR FILING DATE: 2003-02-05
NUMBER OF SEQ ID NOS: 78
SOFTWARE: PatentIn version 3.2
SEQ ID NO 44
LENGTH: 384
TYPE: PRT
ORGANISM: Nicotiana tabacum
PCT-US03-20001-44
```

```
Query Match          44.0% Score 40; DB 1; Length 384;
Best Local Similarity 61.5% Pred. No. 2.1e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
```

```
QY      1 KSPSEQEVTLDGN 13
      |||:|:|
Db      145 KDPALQETILRGN 157
```

```
RESULT 21
```

US-10-469-204-167  
; Sequence 167, Application US/10469204  
; GENERAL INFORMATION:  
; APPLICANT: DSM NV  
; TITLE OF INVENTION: Novel genes encoding novel proleolytic enzymes.  
; FILE REFERENCE: 20095W0  
; CURRENT APPLICATION NUMBER: US/10/469,204  
; CURRENT FILING DATE: 2003-08-25  
; NUMBER OF SEQ ID NOS: 171  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 167  
; LENGTH: 623  
; TYPE: PRT  
; ORGANISM: Aspergillus niger  
US-10-469-204-167

Query Match 44.0%; Score 40; DB 6; Length 623;  
Best Local Similarity 50.0%; Pred. No. 2.8e+02;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 EQQETVLDGNLIIR 17  
||:||||:|  
Db 582 EQREVVLSGDAVVR 595

RESULT 22  
PCT-US03-29089-28  
; Sequence 28, Application PC/TUS0329089  
; GENERAL INFORMATION:  
; APPLICANT: EXELIXIS, INC.  
; TITLE OF INVENTION: MEMS AS MODIFIERS OF BRANCHING MORPHOGENESIS AND METHODS OF USE  
; FILE REFERENCE: EX03-062C-PC  
; CURRENT APPLICATION NUMBER: PCT/US03/29089  
; CURRENT FILING DATE: 2003-09-15  
; PRIOR APPLICATION NUMBER: US 60/411,154  
; PRIOR FILING DATE: 2002-09-16  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 28  
; LENGTH: 1034  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US03-29089-28

Query Match 42.9%; Score 39; DB 1; Length 1034;  
Best Local Similarity 66.7%; Pred. No. 4.9e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KSPQEQETVLDG 12  
:|||||  
Db 722 QSPQEQSTVLRG 733

RESULT 23  
US-60-495-589-224  
; Sequence 224, Application US/60495589  
; GENERAL INFORMATION:  
; APPLICANT: Proguiske-Fox, Ann  
; APPLICANT: Hallman, Jeffrey D.  
; APPLICANT: Handfield, Martin  
; TITLE OF INVENTION: IDENTIFICATION OF PORPHYROMONAS GINGIVALIS VIRULENCE POLYNUCLEOTI  
; FILE REFERENCE: 02-042  
; CURRENT APPLICATION NUMBER: US/60/495,589  
; CURRENT FILING DATE: 2003-08-15  
; NUMBER OF SEQ ID NOS: 354  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 224  
; LENGTH: 1132  
; TYPE: PRT  
; ORGANISM: Porphyromonas gingivalis  
US-60-495-589-224

Query Match 42.9%; Score 39; DB 7; Length 1132;  
Best Local Similarity 50.0%; Pred. No. 5.2e+02;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 SPQEQVLDGNLI 15  
||:||||:|  
Db 403 SPQEQVALDPELL 416

RESULT 24  
PCT-US03-28508-59  
; Sequence 59, Application PC/TUS0328508  
; GENERAL INFORMATION:  
; APPLICANT: Lu, Peter S.  
; APPLICANT: Schweizer, Johannes  
; APPLICANT: Diaz-Sarmiento, Chamorro Somoza  
; APPLICANT: Belmares, Michael P.  
; TITLE OF INVENTION: METHODS OF DIAGNOSING CERVICAL CANCER  
; FILE REFERENCE: VITA-008W0  
; CURRENT APPLICATION NUMBER: PCT/US03/28508  
; CURRENT FILING DATE: 2003-09-09  
; PRIOR APPLICATION NUMBER: 60/409,298  
; PRIOR FILING DATE: 2002-09-09  
; PRIOR APPLICATION NUMBER: 60/450,464  
; PRIOR FILING DATE: 2003-02-27  
; PRIOR APPLICATION NUMBER: 10/630,590  
; PRIOR FILING DATE: 2003-07-29  
; PRIOR APPLICATION NUMBER: 60/490,094  
; PRIOR FILING DATE: 2003-07-25  
; NUMBER OF SEQ ID NOS: 330  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 59  
; LENGTH: 109  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
PCT-US03-28508-59

Query Match 41.8%; Score 38; DB 1; Length 109;  
Best Local Similarity 37.5%; Pred. No. 1.6e+02;  
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 2 SPQEQETVLDGNLIIR 17  
||:||||:|  
Db 55 SPQEQSGIDQDITVK 70

RESULT 25  
US-10-425-114A-64443  
; Sequence 64443, Application US/10425114A  
; GENERAL INFORMATION:  
; APPLICANT: Lin, Jindong  
; APPLICANT: Zhou, Yinhua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114A  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 64443  
; LENGTH: 209  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3959-001-A11\_F11.dep  
US-10-425-114A-64443

Query Match 41.8%; Score 38; DB 6; Length 209;  
Best Local Similarity 58.3%; Pred. No. 2.4e+02;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```
OY      4 EQETVLDGNI 15
      :|:|:|:|
Db      159 DQPNVLDGNI 170

RESULT 26
US-09-897-516A-5308
; Sequence 5308, Application US/09897516A
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldan, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Thomas M.
; APPLICANT: Malvar, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215,161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8415
; SEQ ID NO 5308
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-09-897-516A-5308

Query Match      41.8%; Score 38; DB 5; Length 294;
Best Local Similarity 77.8%; Pred. No. 3e+02;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY      2 SPEQETVL 10
      :|:|:|:|
Db      113 APEQDPTVL 121

RESULT 27
US-60-495-114-2249
; Sequence 2249, Application US/60495114
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CLO01480
; CURRENT APPLICATION NUMBER: US/60/495,114
; CURRENT FILING DATE: 2003-08-15
; NUMBER OF SEQ ID NOS: 91238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2249
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-495-114-2249

Query Match      41.8%; Score 38; DB 7; Length 353;
Best Local Similarity 63.6%; Pred. No. 3.4e+02;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      6 QETVLDGNI 16
      :|:|:|:|
Db      158 QDTINDGNI 168

RESULT 28
US-10-644-807-430
; Sequence 430, Application US/10644807
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.

; TITLE OF INVENTION: 83 Human Secreted Proteins
; FILE REFERENCE: PS735
; CURRENT APPLICATION NUMBER: US/10/644,807
; CURRENT FILING DATE: 2003-08-21
; PRIOR APPLICATION NUMBER: PCT/US02/05064
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: 60/270,658
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/304,444
; PRIOR FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 445
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 430
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (189)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (253)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-644-807-430

Query Match      41.8%; Score 38; DB 6; Length 356;
Best Local Similarity 44.0%; Pred. No. 3.4e+02;
Matches 11; Conservative 4; Mismatches 2; Indels 8; Gaps 2;

OY      2 SPEQET----VL---DGNLIIRY 18
      :|:|:|:|:|:|:|:|:|:|
Db      77 APEQFTRVGVQVLDKRDGSRIVRY 101

RESULT 29
PCT-US03-26780-1271
; Sequence 1271, Application PC/TUS0326780
; GENERAL INFORMATION:
; APPLICANT: FIVEPRIME THERAPEUTICS, INC.
; TITLE OF INVENTION: HUMAN POLYPEPTIDES ENCODED BY POLYNUCLEOTIDES AND METHODS OF
; TITLE OF INVENTION: THEIR USE
; FILE REFERENCE: 08940.0014-00304
; CURRENT APPLICATION NUMBER: PCT/US03/26780
; CURRENT FILING DATE: 2003-08-28
; PRIOR APPLICATION NUMBER: 60/406,616
; PRIOR FILING DATE: 2002-08-29
; PRIOR APPLICATION NUMBER: 60/406,579
; PRIOR FILING DATE: 2002-08-29
; PRIOR APPLICATION NUMBER: 60/406,655
; PRIOR FILING DATE: 2002-08-29
; PRIOR APPLICATION NUMBER: 60/406,642
; PRIOR FILING DATE: 2002-08-29
; PRIOR APPLICATION NUMBER: 60/406,640
; PRIOR FILING DATE: 2002-08-29
; PRIOR APPLICATION NUMBER: 60/406,588
; PRIOR FILING DATE: 2002-08-29
; PRIOR APPLICATION NUMBER: 60/406,576
; PRIOR FILING DATE: 2002-08-29
; PRIOR APPLICATION NUMBER: 60/406,646
; PRIOR FILING DATE: 2002-08-29
; PRIOR APPLICATION NUMBER: 60/406,666
; PRIOR FILING DATE: 2002-08-29
; PRIOR APPLICATION NUMBER: 60/406,653
; PRIOR FILING DATE: 2002-08-29
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 3700
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1271
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US03-26780-1271
```



Query Match 41.8%; Score 38; DB 1; Length 359;  
Best Local Similarity 44.0%; Pred. No. 3.4e+02;  
Matches 11; Conservative 4; Mismatches 2; Indels 8; Gaps 2;

Oy 2 SPEGQET-----VL---DGNLIIRY 18  
Db 77 APEQFTRVGVQVLDKRDGSEFIVY 101

## RESULT 30

US-10-644-807-344  
Sequence 344, Application US/10644807  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 83 Human Secreted Proteins  
FILE REFERENCE: PS735  
CURRENT APPLICATION NUMBER: US/10/644, 807  
PRIOR FILING DATE: 2003-08-21  
PRIOR APPLICATION NUMBER: PCT/US02/05064  
PRIOR FILING DATE: 2002-02-21  
PRIOR APPLICATION NUMBER: 60/270,658  
PRIOR FILING DATE: 2001-02-23,204,444  
PRIOR APPLICATION NUMBER: 60/304,444  
PRIOR FILING DATE: 2001-07-12  
NUMBER OF SEQ ID NOS: 445  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 344  
LENGTH: 365  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (189)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (253)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (365)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-10-644-807-344

Query Match 41.8%; Score 38; DB 6; Length 365;  
Best Local Similarity 44.0%; Pred. No. 3.4e+02;  
Matches 11; Conservative 4; Mismatches 2; Indels 8; Gaps 2;

Oy 2 SPEGQET-----VL---DGNLIIRY 18  
Db 77 APEQFTRVGVQVLDKRDGSEFIVY 101

## RESULT 31

US-10-425-114A-62212  
Sequence 62212, Application US/10425114A  
GENERAL INFORMATION:  
APPLICANT: Liu, Jindong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E  
APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53113)B  
CURRENT APPLICATION NUMBER: US/10/425, 114A  
PRIOR FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 62212  
LENGTH: 367  
TYPE: PRT

ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: LIB3245-311-F7\_FLI pep  
US-10-425-114A-62212

Query Match 41.8%; Score 38; DB 6; Length 367;  
Best Local Similarity 58.3%; Pred. No. 3.4e+02;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 2 SPEGQETVLDGN 13  
Db 33 TPENFDIVLDGS 44

## RESULT 32

US-60-485-450-1641  
Sequence 1641, Application US/60485450  
GENERAL INFORMATION:  
APPLICANT: CARGILL, Michele  
APPLICANT: CHANG, Sheng-Yung  
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
TITLE OF INVENTION: RESPONSE TO INTERFERON TREATMENT IN HEPATITIS C  
TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES  
FILE REFERENCE: CL001470  
CURRENT APPLICATION NUMBER: US/60/485, 450  
PRIOR FILING DATE: 2003-07-09  
NUMBER OF SEQ ID NOS: 47859  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1641  
LENGTH: 377  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-60-485-450-1641

Query Match 41.8%; Score 38; DB 7; Length 377;  
Best Local Similarity 53.8%; Pred. No. 3.5e+02;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Oy 4 BQETVLDGNLI 16  
Db 149 EREERVMDGLVI 161

## RESULT 33

PCT-US03-20001-32  
Sequence 32, Application PC/TUS0320001  
GENERAL INFORMATION:  
APPLICANT: Boyes, Douglas  
APPLICANT: Davis, Keith  
APPLICANT: Jones, Alan  
APPLICANT: Ullah, Hemayet  
APPLICANT: Chen, Jin-Gui  
APPLICANT: Mulpuri, Rao  
APPLICANT: Chatterjee, Anil  
TITLE OF INVENTION: METHODS FOR IMPROVING PLANT AGRONOMICAL TRAITS BY ALTERING THE  
TITLE OF INVENTION: EXPRESSION OR ACTIVITY OF PLANT G-PROTEIN ALPHA AND BETA SUBUNIT.  
FILE REFERENCE: 215505  
CURRENT APPLICATION NUMBER: PCT/US03/20001  
PRIOR FILING DATE: 2003-06-24, 730  
PRIOR APPLICATION NUMBER: 60/392,730  
PRIOR FILING DATE: 2002-06-28  
PRIOR APPLICATION NUMBER: 60/445,208  
PRIOR FILING DATE: 2003-02-05  
NUMBER OF SEQ ID NOS: 78  
SOFTWARE: PatentIn Version 3.2  
SEQ ID NO 32  
LENGTH: 384  
TYPE: PRT  
ORGANISM: Solanum tuberosum  
PCT-US03-20001-32

Query Match 41.8%; Score 38; DB 1; Length 384;  
 Best Local Similarity 61.5%; Pred. No. 3.5e+02;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 KSPEQETVLDGN 13  
 |||:|  
 Db 145 KDPALQETLRLGN 157

RESULT 34  
 PCT-US03-20001-51  
 ; Sequence 51, Application PC/TUS0320001  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boyes, Douglas  
 ; APPLICANT: Davis, Keith  
 ; APPLICANT: Jones, Alan  
 ; APPLICANT: Ullah, Hemayet  
 ; APPLICANT: Chen, Jin-Gui  
 ; APPLICANT: Mulpuri, Rao  
 ; APPLICANT: Chatterjee, Ani  
 ; APPLICANT: Ward, Mary  
 ; TITLE OF INVENTION: METHODS FOR IMPROVING PLANT AGRONOMICAL TRAITS BY ALTERING THE  
 ; TITLE OF INVENTION: EXPRESSION OR ACTIVITY OF PLANT G-PROTEIN ALPHA AND BETA SUBUNIT  
 ; FILE REFERENCE: 2155US  
 ; CURRENT APPLICATION NUMBER: PCT/US03/20001  
 ; CURRENT FILING DATE: 2003-06-24  
 ; PRIOR APPLICATION NUMBER: 60/392,730  
 ; PRIOR FILING DATE: 2002-06-28  
 ; PRIOR APPLICATION NUMBER: 60/445,208  
 ; PRIOR FILING DATE: 2003-02-05  
 ; NUMBER OF SEQ ID NOS: 78  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 51  
 ; LENGTH: 384  
 ; TYPE: PRT  
 ; ORGANISM: Lycopersicon esculentum  
 PCT-US03-20001-51

Query Match 41.8%; Score 38; DB 1; Length 384;  
 Best Local Similarity 61.5%; Pred. No. 3.5e+02;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 KSPEQETVLDGN 13  
 |||:|  
 Db 145 KDPALQETLRLGN 157

RESULT 35  
 PCT-US03-20001-28  
 ; Sequence 28, Application PC/TUS0320001  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boyes, Douglas  
 ; APPLICANT: Davis, Keith  
 ; APPLICANT: Jones, Alan  
 ; APPLICANT: Ullah, Hemayet  
 ; APPLICANT: Chen, Jin-Gui  
 ; APPLICANT: Mulpuri, Rao  
 ; APPLICANT: Chatterjee, Ani  
 ; APPLICANT: Ward, Mary  
 ; TITLE OF INVENTION: METHODS FOR IMPROVING PLANT AGRONOMICAL TRAITS BY ALTERING THE  
 ; TITLE OF INVENTION: EXPRESSION OR ACTIVITY OF PLANT G-PROTEIN ALPHA AND BETA SUBUNIT  
 ; FILE REFERENCE: 2155US  
 ; CURRENT APPLICATION NUMBER: PCT/US03/20001  
 ; CURRENT FILING DATE: 2003-06-24  
 ; PRIOR APPLICATION NUMBER: 60/392,730  
 ; PRIOR FILING DATE: 2002-06-28  
 ; PRIOR APPLICATION NUMBER: 60/445,208  
 ; PRIOR FILING DATE: 2003-02-05  
 ; NUMBER OF SEQ ID NOS: 78  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 28  
 ; LENGTH: 385  
 ; TYPE: PRT

; ORGANISM: Solanum tuberosum  
 PCT-US03-20001-28

Query Match 41.8%; Score 38; DB 1; Length 385;  
 Best Local Similarity 61.5%; Pred. No. 3.5e+02;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 KSPEQETVLDGN 13  
 |||:|  
 Db 146 KDPALQETLRLGN 158

RESULT 36  
 PCT-US03-20001-30  
 ; Sequence 30, Application PC/TUS0320001  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boyes, Douglas  
 ; APPLICANT: Davis, Keith  
 ; APPLICANT: Jones, Alan  
 ; APPLICANT: Ullah, Hemayet  
 ; APPLICANT: Chen, Jin-Gui  
 ; APPLICANT: Mulpuri, Rao  
 ; APPLICANT: Chatterjee, Ani  
 ; APPLICANT: Ward, Mary  
 ; TITLE OF INVENTION: METHODS FOR IMPROVING PLANT AGRONOMICAL TRAITS BY ALTERING THE  
 ; TITLE OF INVENTION: EXPRESSION OR ACTIVITY OF PLANT G-PROTEIN ALPHA AND BETA SUBUNIT  
 ; FILE REFERENCE: 2155US  
 ; CURRENT APPLICATION NUMBER: PCT/US03/20001  
 ; CURRENT FILING DATE: 2003-06-24  
 ; PRIOR APPLICATION NUMBER: 60/392,730  
 ; PRIOR FILING DATE: 2002-06-28  
 ; PRIOR APPLICATION NUMBER: 60/445,208  
 ; PRIOR FILING DATE: 2003-02-05  
 ; NUMBER OF SEQ ID NOS: 78  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 30  
 ; LENGTH: 392  
 ; TYPE: PRT  
 ; ORGANISM: Solanum tuberosum  
 PCT-US03-20001-30

Query Match 41.8%; Score 38; DB 1; Length 392;  
 Best Local Similarity 61.5%; Pred. No. 3.6e+02;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 KSPEQETVLDGN 13  
 |||:|  
 Db 153 KDPALQETLRLGN 165

RESULT 37  
 US-60-485-450-1640  
 ; Sequence 1640, Application US/60485450  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CARGILL, Michele  
 ; APPLICANT: CHANG, Sheng-Yung  
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
 ; TITLE OF INVENTION: RESPONSE TO INTERPERON TREATMENT IN HEPATITIS C  
 ; TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES  
 ; TITLE OF INVENTION: THEREOF  
 ; FILE REFERENCE: CL001470  
 ; CURRENT APPLICATION NUMBER: US/60/485,450  
 ; CURRENT FILING DATE: 2003-07-09  
 ; NUMBER OF SEQ ID NOS: 47859  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1640  
 ; LENGTH: 393  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-60-485-450-1640

Query Match 41.8%; Score 38; DB 7; Length 393;  
 Best Local Similarity 53.8%; Pred. No. 3.6e+02;

Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 QY 4 EQQETVLDGNLII 16  
 Db 165 EREERVMDGLIVI 177

RESULT 38  
 US-60-485-450-1638  
 ; Sequence 1638, Application US/60485450  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CARGILL, Michele  
 ; APPLICANT: CHANG, Sheng-Yung  
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
 ; TITLE OF INVENTION: RESPONSE TO INTERFERON TREATMENT IN HEPATITIS C  
 ; TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES  
 ; FILE REFERENCE: CL001470  
 ; CURRENT APPLICATION NUMBER: US/60/485,450  
 ; CURRENT FILING DATE: 2003-07-09  
 ; NUMBER OF SEQ ID NOS: 47859  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1638  
 ; LENGTH: 423  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-60-485-450-1638

Query Match 41.8%; Score 38; DB 7; Length 423;  
 Best Local Similarity 53.8%; Pred. No. 3.8e+02;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 EQQETVLDGNLII 16  
 Db 195 EREERVMDGLIVI 207

RESULT 39  
 US-60-485-450-1639  
 ; Sequence 1639, Application US/60485450  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CARGILL, Michele  
 ; APPLICANT: CHANG, Sheng-Yung  
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
 ; TITLE OF INVENTION: RESPONSE TO INTERFERON TREATMENT IN HEPATITIS C  
 ; TITLE OF INVENTION: VIRUS-INFECTED SUBJECTS, METHODS OF DETECTION AND USES  
 ; FILE REFERENCE: CL001470  
 ; CURRENT APPLICATION NUMBER: US/60/485,450  
 ; CURRENT FILING DATE: 2003-07-09  
 ; NUMBER OF SEQ ID NOS: 47859  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1639  
 ; LENGTH: 439  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-60-485-450-1639

Query Match 41.8%; Score 38; DB 7; Length 439;  
 Best Local Similarity 53.8%; Pred. No. 3.8e+02;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 4 EQQETVLDGNLII 16  
 Db 211 EREERVMDGLIVI 223

RESULT 40  
 US-60-495-114-1338  
 ; Sequence 1338, Application US/60495114  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CARGILL, Michele  
 ; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES

; TITLE OF INVENTION: ENCODING HUMAN PROTEASE PROTEINS, METHODS OF DETECTION AND  
 ; TITLE OF INVENTION: USES THEREOF  
 ; FILE REFERENCE: CL001480  
 ; CURRENT APPLICATION NUMBER: US/60/495,114  
 ; CURRENT FILING DATE: 2003-08-15  
 ; NUMBER OF SEQ ID NOS: 91238  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1338  
 ; LENGTH: 452  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-60-495-114-1338

Query Match 41.8%; Score 38; DB 7; Length 452;  
 Best Local Similarity 37.5%; Pred. No. 3.9e+02;  
 Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 2 SPEQETVLDGNLII 17  
 Db 394 SPQORGGIODGDIIVK 409

Search completed: October 10, 2003, 17:33:31  
 Job time : 25 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: October 10, 2003, 17:11:49 ; Search time 83 Seconds

(without alignments)  
34.423 Million cell updates/sec

Title: US-09-991-795-1

Perfect score: 91

Sequence: 1 KSPGQETVDGNLIRY 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	91	100.0	339	AAU32845	Novel human secret
2	91	100.0	930	ABO9708	Sequence of H4p he
3	91	100.0	930	ABO9709	Amino acid sequenc
4	85	94.5	123	AAU32845	Novel human secret
5	85	94.5	921	ABO9711	Sequence of H4p he
6	79	86.8	16	ABBS2029	Human APT-57 ctyp
7	79	86.8	16	ABG78741	Multiple sclerosis
8	79	86.8	16	ABP57165	Breast cancer asso
9	57	62.6	932	ABO9706	Sequence of H4p he

10	57	62.6	933	ABO9707	Sequence of H4p he
11	46	50.5	179	ABP25414	Streptococcus poly
12	46	50.5	179	ABP29767	Streptococcus poly
13	45	49.5	452	ABBS2562	Herbically activ
14	44	48.4	409	AAK41227	910 SLG protein.
15	43	47.3	650	AAK5180	Cellulose synthase
16	43	47.3	858	ABP27072	Streptococcus poly
17	43	47.3	946	AAK29100	Human inter-alpha-
18	42	46.2	28	ABP70358	Tryptic peptide of
19	42	46.2	419	AAW10977	dihydrofolate redu
20	42	46.2	419	AAW11865	DHFR/Polypeptide B
21	42	46.2	544	AAW10975	Chlamydia pneumoni
22	42	46.2	544	AAW11863	Polypeptide B. Ch
23	42	46.2	544	AAK34735	Amino acid sequenc
24	42	46.2	544	AAK11757	Chlamydia pneumoni
25	42	46.2	544	ABBS4272	Chlamydia pneumoni
26	42	46.2	652	AAK13000	Corn poly (A) bind
27	42	46.2	652	AAK13002	glycine max poly (
28	42	46.2	704	AAW10976	dihydrofolate redu
29	42	46.2	704	AAW11864	DHFR/Polypeptide B
30	42	46.2	858	AAK53404	S-Locus receptor (
31	42	46.2	858	AAW49080	Brassica sp. S-rec
32	42	46.2	2515	ABBS4427	Drosophila melanog
33	41.5	45.6	346	AAK2710	S. epidermidis ope
34	41.5	45.6	353	ABP39795	Staphylococcus epi
35	41.5	45.6	1254	AAK79929	CGGD-4, Incyte ID
36	41	45.1	28	ABP70321	Tryptic peptide of
37	41	45.1	140	AAK10209	Human bone marrow
38	41	45.1	189	AAU18354	Human endocrine po
39	41	45.1	258	AAK37099	Amino acid sequenc
40	41	45.1	276	ABBS48710	Listeria monocytog
41	41	45.1	436	ABBS4533	R. solivus RPS 5.2
42	41	45.1	444	ABP73867	Candida albicans e
43	41	45.1	544	AAK3357	HypB protein. Chl
44	41	45.1	612	AAU34514	E. coli cellular p
45	41	45.1	628	ABG24763	Novel human diagno

## ALIGNMENTS

RESULT 1	AAU32845
ID	AAU32845 standard; Protein: 339 AA.
XX	AAU32845;
AC	18-DEC-2001 (first entry)
XX	
DT	Novel human secreted protein #3336.
XX	
DE	Human; vaccination; gene therapy; nutritional supplement;
XX	stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW	immune suppression; immune stimulation; anti-inflammatory; leukaemia.
KM	
XX	
OS	Homo sapiens.
XX	
PN	WO200179449-A2.
XX	
PD	25-OCT-2001.
XX	
PF	16-APR-2001; 2001WO-US08656.
XX	
PR	18-APR-2000; 2000US-0552929.
XX	
PR	26-JAN-2001; 2001US-0770160.
XX	
PA	(HYSB-) HYSBQ INC.
XX	
PI	Tang YT, Liu C, Drmanac RT;
XX	
DR	WPI; 2001-611725/70.
XX	
PT	Nucleic acids encoding a range of human polypeptides, useful in genetic

PT vaccination, testing and therapy -  
XX  
PS  
XX Claim 20; Page 678; 7655p; English.

CC The invention relates to novel human secreted polypeptides. The  
CC polypeptides and antibodies to the polypeptides are useful for  
CC determining the presence of or predisposition to a disease associated  
CC with altered levels of polypeptide. The polypeptides are also useful for  
CC identifying agents (agonists and antagonists) that bind to them. Cells  
CC expressing the proteins are useful for identifying a therapeutic agent  
CC for use in treatment of a pathology related to aberrant expression or  
CC physiological interactions of the polypeptide. Vectors comprising  
CC the nucleic acid encoding the polypeptides and cells genetically  
CC engineered to express them are also useful for producing the proteins.  
CC The proteins are useful in genetic vaccination, testing and  
CC therapy, and can be used as nutritional supplements. They may be used to  
CC increase stem cell proliferation; to regulate haemopoiesis; and in  
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;  
CC immune suppression and/or stimulation; as anti-inflammatory agents; and  
CC in treatment of leukaemias. ANU29510-ANU3304 represent the amino acid  
CC sequences of novel human secreted proteins of the invention.

Query Match	100.0%	Score 91	DB 22	Length 339
Best Local Similarity	100.0%	Pred. No.	1.4e-07	
Matches 18	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY 1 KSPQEQETVLDDGNLIIRY 18  
124 KSPQEQETVLDDGNLIIRY 141

RESULT 2	
ABB09708	
ID	ABB09708 standard; Protein; 930 AA

AC	ABB09708;
XX	
DT	11-JUN-2002 (first entry)

DE Sequence of H4P heavy chain of inter alpha trypsin inhibitor.

KM MAY-1; blood transfusion; immunosuppressive; mixed lymphocyte reaction;  
KM MLE; autoimmune disease; allergic disease; organ rejection;  
KM organ transplantation; rheumatism; psoriasis; bronchial asthma;  
KM allergic rhinitis; allergic dermatitis; pollinosis; HAP heavy chain;  
KM inter alpha trypsin inhibitor.

OS Homo sapiens.

PN WO200212495-A1.

PD 14-FEB-2002.

PF 01-AUG-2001; 2001WO-JP066620.

PR 09-AUG-2000; 2000JP-0241169.

PA (MARU-) MARUHO KK.

PI Uchida H, Tanaka H, Kitch Y, Fujimura A, Kobayashi E, ...

DR WPI; 2002-217191/27.

**XX**

PT Rat or human protein MAY-1 induced by homogeneous blood transfusion and  
PT exhibiting immunosuppressive activity in a homogeneous mixed lymphocyte  
PT reaction, for providing drug compositions to treat e.g. autoimmune  
PT diseases -

PS Disclosure; Page 55-59; 85pp; Japanese.

CC The present sequence represents the human H4P heavy chain of inter  
CC alpha trypsin inhibitor protein. The specification describes MAY-1  
CC protein, which is induced by homogeneous blood transfusion. MAY-1  
CC exhibits an immunosuppressive activity in a homogeneous mixed  
CC lymphocyte reaction (MLR). The MAY-1 protein can formulated into drug  
CC compositions for the prevention or treatment of autoimmune diseases,  
CC allergic diseases, or rejection reaction during organ transplantation,  
CC e.g. rheumatism, psoriasis, bronchial asthma, allergic rhinitis,  
CC allergic dermatitis and pollinosis.  
CX  
XX Sequence 930 Aa;

Query Match	100.0%	Score 91;	DB 23;	Length 930;
Best Local Similarity	100.0%;	Pred. No. 5e-07;		
Matches 18;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 1 KSPEQOETVL DGNLIRY 18  
|||  
Db 224 KSPEQOETVL DGNLIRY 241

150

RESULT 3	
ABB09709	
ID	ABB09709 standard; Protein; 930 AA.

AC ABB09709;

DT 11-JUN-2002 (first entry)

DE Amino acid sequence of a human PK-120 polypeptide

KW MAX-1; blood transfusion; immunosuppressive; mixed lymphocyte reaction;  
KW MMR; autoimmune disease; allergic disease; organ rejection  
KW organ transplantation; rheumatism; psoriasis; bronchial asthma;  
KW allergic rhinitis; allergic dermatitis; pollinosis; PK-120; ss.

OS Homo sapiens.

PN WO200212495-A1

PD 14-FEB-2002.

PF 01-AUG-2001; 2001WO-JP066620.

PR 09-AUG-2000; 2000JP-0241169.

PA (MARU-) MARUHO KK

PI Uchida H, Tanaka H, Kitch Y, Fujimura A, Kobayashi E,

DR WPI; 2002-217191/27.

—

PT Rat or human protein MAY-1 induced by homogeneous blood transfusion and  
PT exhibiting immunosuppressive activity in a homogeneous mixed lymphocyte  
PT reaction, for providing drug compositions to treat e.g. autoimmune  
PT diseases -

PS Disclosure; Page 62-66; 85pp; Japanese.

CC The present sequence represents a human PK-120 polypeptide. The  
CC specification describes MAY-1 protein, which is induced by  
CC homogeneous blood transfusion. MAY-1 exhibits an immunosuppressive  
CC activity in a homogeneous mixed lymphocyte reaction (MLR). The MAY-1  
CC protein can formulated into drug compositions for the prevention or  
CC treatment of autoimmune diseases, allergic diseases, or rejection  
CC reaction during organ transplantation, e.g. Rheumatism, psoriasis,  
CC bronchial asthma, allergic rhinitis, allergic dermatitis and pollinosis

**SQ Sequence 930 AA;**

Query Match	100.0%;	Score 91;	DB 23;	Length 930;
Best Local Similarity	100.0%;	Pred. No. 5e-07;		

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 KSPQOETVLDGNLIIRY 18  
 |||||  
 Db 224 KSPQOETVLDGNLIIRY 241  
 |||||

20

RESULT 4  
 ID AAU32844 standard; Protein: 123 AA.

XX AAU32844;  
 AC AAU32844;  
 DT 18-DEC-2001 (first entry)  
 DE Novel human secreted protein #3335.

XX Human; vaccination; gene therapy; nutritional supplement;  
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;  
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

OS Homo sapiens.  
 PN WO200179449-A2.  
 PD 25-OCT-2001.

PF 16-APR-2001; 2001WO-US08656.  
 PR 18-APR-2000; 2000US-0552929.  
 PR 26-JAN-2001; 2001US-0770160.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

WP1; 2001-611725/70.

PT Nucleic acid encoding a range of human polypeptides, useful in genetic  
 PT vaccination, testing and therapy -  
 PS Claim 20; Page 676-678; 765pp; English.

CC The invention relates to novel human secreted polypeptides. The  
 CC polypeptides and antibodies to the polypeptides are useful for  
 CC determining the presence of or predisposition to a disease associated  
 CC with altered levels of polypeptide. The polypeptides are also useful for  
 CC identifying agents (agonists and antagonists) that bind to them. Cells  
 CC expressing the proteins are useful for identifying a therapeutic agent  
 CC for use in treatment of a pathology related to aberrant expression or  
 CC physiological interactions of the polypeptide. Vectors comprising  
 CC the nucleic acids encoding the polypeptides and cells genetically  
 CC engineered to express them are also useful for producing the proteins.  
 CC The proteins are useful in genetic vaccination, testing and  
 CC therapy, and can be used as nutritional supplements. They may be used to  
 CC increase stem cell proliferation; to regulate haematopoiesis; and in  
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;  
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and  
 CC in treatment of leukaemias. AUA29510-AUA3304 represent the amino acid  
 CC sequences of novel human secreted proteins of the invention.

SO Sequence 123 AA;

Query Match 94.5%; Score 86; DB 22; Length 123;  
 Best Local Similarity 94.4%; Pred. No. 3.1e-07;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 KSPQOETVLDGNLIIRY 18  
 |||||  
 Db 53 KSPQOETVLDGNLIIRY 70  
 |||||

RESULT 5

ABB09711  
 ID ABB09711 standard; Protein: 921 AA.

XX ABB09711;  
 AC ABB09711;  
 DT 11-JUN-2002 (first entry)

DE Sequence of H4P heavy chain of inter-alpha-inhibitor protein.

KW MAY-1, blood transfusion; immunosuppressive; mixed lymphocyte reaction;  
 KW MMR; autoimmune disease; allergic disease; organ rejection;  
 KW organ transplantation; rheumatism; psoriasis; bronchial asthma;  
 KW allergic rhinitis; allergic dermatitis; pollinosis; H4P heavy chain;  
 KW inter-alpha-inhibitor protein.

OS Sgs sp.  
 PN WO200212495-A1.  
 PD 14-FEB-2002.

PF 01-AUG-2001; 2001WO-JP06620.  
 PR 09-AUG-2000; 2000JP-0241169.

PA (MARU-) MARUHO KK.

PI Uchida H, Tanaka H, Kitch Y, Fujimura A, Kobayashi E;

WP1; 2002-217191/27.  
 DR N-PSDB; ABL41975.

PT Rat or human protein MAY-1 induced by homogeneous blood transfusion and  
 PT exhibiting immunosuppressive activity in a homogeneous mixed lymphocyte  
 PT reaction, for providing drug compositions to treat e.g. autoimmune  
 PT diseases -

PS Disclosure; Page 71-75; 85pp; Japanese.

CC The present sequence represents the hog H4P heavy chain of  
 CC inter-alpha-inhibitor protein. The specification describes MAY-1  
 CC protein, which is induced by homogeneous blood transfusion. MAY-1  
 CC exhibits an immunosuppressive activity in a homogeneous mixed  
 CC lymphocyte reaction (MLR). The MAY-1 protein can formulated into drug  
 CC compositions for the prevention or treatment of autoimmune diseases,  
 CC allergic diseases, or rejection reaction during organ transplantation,  
 CC e.g. Rheumatism, psoriasis, bronchial asthma, allergic rhinitis,  
 CC allergic dermatitis and pollinosis.

SO Sequence 921 AA;

Query Match 94.5%; Score 86; DB 23; Length 921;  
 Best Local Similarity 88.9%; Pred. No. 3.8e-06;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 KSPQOETVLDGNLIIRY 18  
 |||||  
 Db 222 KSPQOETVLDGNLIIRY 239  
 |||||

RESULT 6

ABB52029  
 ID ABB52029 standard; Peptide: 16 AA.

XX ABB52029;  
 AC ABB52029;  
 DT 08-FEB-2002 (first entry)

DE Human API-57 tryptic digest peptide #2.  
 KW Human; neuroprotective; nootropic; gene therapy; vaccine;  
 KW Alzheimer's disease; Alzheimer's Disease-Associated Feature; AF;  
 KW Alzheimer's Disease-Associated Protein Isoform; API; tryptic digest;

KM Expression Reference Protein Isoform; ERPI; proteolysis.  
 XX Homo sapiens.  
 OS  
 XX MO200175454-A2.  
 PN  
 XX 11-OCT-2001.  
 PD  
 XX 03-APR-2001; 2001WO-US10908.  
 PF  
 XX 03-APR-2000; 2000US-194504P.  
 PR  
 XX 28-NOV-2000; 2000US-253647P.  
 PP  
 XX (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
 PA (PRIZ ) PRIZER INC.  
 PI Durham XL, Friedman DL, Herath HMAc, Kimmel LH, Parekh RB;  
 PI Potter DM, Rohlf C, Silber BM, Sciger TR, Sunderland PR;  
 PI Townsend RR, White F, Williams SA;  
 XX WPI; 2001-639384/73.  
 DR  
 XX Screening for Alzheimer's disease in a mammal, by making  
 PT two-dimensional array of a feature whose relative abundance correlates  
 PT with disease, and comparing with abundance of the feature in samples of  
 PT healthy persons -  
 PS Example; Page 26; 162pp; English.  
 XX The invention relates to methods for the screening, diagnosis and  
 CC prognosis of Alzheimer's disease. The methods involve the detection  
 CC of Alzheimer's Disease-Associated Features (AFs) and Alzheimer's  
 CC Disease-Associated Protein Isoforms (APis) in cerebrospinal fluid,  
 CC serum or plasma. The abundance of the APis and APis is then  
 CC normalised to an Expression Reference Protein Isoform (ERPI) in  
 CC order to determine whether a patient is suffering from, or has  
 CC a predisposition to, Alzheimer's Disease. The relative abundance of  
 CC the APis and APis correlates with the severity of Alzheimer's Disease.  
 CC The present sequence is a peptide produced from an APi by proteolysis.  
 CC  
 SQ Sequence 16 AA;  
 Query Match 86.8%; Score 79; DB 22; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-07;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 SPEQETVLDGNLIR 17  
 DB 1 SPEQETVLDGNLIR 16  
 RESULT 7  
 ABG78741  
 ID ABG78741 standard; Peptide; 16 AA.  
 AC  
 XX ABG78741;  
 XX  
 DT 29-NOV-2002 (first entry)  
 DE Multiple sclerosis associated feature (MSF) tryptic digest peptide #229.  
 XX  
 XX Multiple sclerosis; MS; multiple sclerosis associated feature; MSaF;  
 KW human; multiple sclerosis-associated protein isoform; MSPI;  
 KW antiinflammatory; neuroprotective.  
 OS Homo sapiens.  
 OS  
 XX MO200259604-A2.  
 PN  
 XX 01-AUG-2002.  
 PD  
 XX 25-JAN-2002; 2002WO-GB00330.  
 PP  
 XX

PR 26-JAN-2001; 2001US-264404P.  
 PR 20-NOV-2001; 2001US-331647P.  
 PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
 PI Herath HMAc, Parekh RB, Rohlf C;  
 PI WPI; 2002-599812/64.  
 DR  
 XX Screening or diagnosing multiple sclerosis (MS), useful for e.g.  
 PT determining the stage or severity of MS, comprises detecting the  
 PT presence of MS-associated features or protein isoforms by 2-dimensional  
 PT electrophoresis -  
 PS Disclosure; Page 27; 128pp; English.  
 XX This invention relates to a novel method for screening or diagnosing  
 CC multiple sclerosis (MS) in a subject to determine the stage or severity  
 CC of MS, to identify a subject at risk of developing MS or to monitor the  
 CC effect of a therapy administered. The method comprises analysing a  
 CC sample body fluid from the subject by two-dimensional electrophoresis  
 CC and detecting the presence of multiple sclerosis-associated features  
 CC (MSFs), or multiple sclerosis-associated protein isoforms (MSPIs).  
 CC The MSFs of the invention correspond to spots identified on a 2D gel  
 CC these proteins may have antiinflammatory or neuroprotective activity.  
 CC The methods of the invention and the compositions are useful for  
 CC clinical screening, diagnosis and treatment of MS, for monitoring the  
 CC effectiveness of MS treatment, for selecting participants in clinical  
 CC trials, for identifying patients most likely to respond to a particular  
 CC therapeutic treatment and for screening and developing drugs for  
 CC treatment of MS. Agents that modulate the expression or activity of an  
 CC MSPI are useful for treating MS, for preventing or delaying the onset or  
 CC development of MS, to prevent or delay the progression of MS, or to  
 CC ameliorate the symptoms MS. Nucleic acids comprising a sequence encoding  
 CC an MSPI, MSPI-related polypeptide, or their fragments are useful for  
 CC promoting MSPI function by gene therapy. The present sequence represents  
 CC a human multiple sclerosis associated feature tryptic digest  
 CC peptide of the invention.  
 CC  
 SQ Sequence 16 AA;  
 Query Match 86.8%; Score 79; DB 23; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-07;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 SPEQETVLDGNLIR 17  
 DB 1 SPEQETVLDGNLIR 16  
 RESULT 8  
 ABP57165  
 ID ABP57165 standard; Peptide; 16 AA.  
 AC  
 XX ABP57165;  
 XX  
 DT 16-APR-2003 (first entry)  
 DE Breast cancer associated tryptic digest SEQ ID NO:116.  
 XX  
 XX Breast cancer associated feature; BF; BPI; breast cancer; diagnosis;  
 KW breast cancer associated protein isoform; cytosolic; gene therapy.  
 OS Homo sapiens.  
 OS Synthetic.  
 XX MO200288750-A2.  
 PN  
 XX 07-NOV-2002.  
 PD  
 XX 02-MAY-2002; 2002WO-GB02022.  
 PP  
 XX 02-MAY-2001; 2001GB-0010790.  
 PR



PR 27-JUN-2001; 2001GB-0018385.  
 PR 14-AUG-2001; 2001GB-0019791.  
 PR 16-AUG-2001; 2001GB-0020045.  
 PR 22-NOV-2001; 2001GB-0028062.  
 XX  
 PA (OXFO-) OXFORD GLYCOSCIENCES UK LTD.  
 XX  
 PI Herath HMAC;  
 XX WPI; 2003-175048/17.  
 XX  
 PT Screening, diagnosing or determining the stage or severity of breast  
 PT cancer, comprises analyzing and quantitatively detecting Breast  
 PT Cancer-Associated Features or Breast Cancer-Associated Protein Isoforms  
 PT in a biological sample -  
 XX  
 PS Disclosure; Page 18; 88pp; English.  
 XX  
 CC The present invention describes a method for screening, diagnosing or  
 CC determining the stage or severity of breast cancer, identifying a subject  
 CC at risk of developing breast cancer, or monitoring the effect of therapy  
 CC administered to a subject with breast cancer, by generating a  
 CC two-dimensional array of features comprising breast cancer-associated  
 CC features (BFS), or quantitatively detecting breast cancer-associated  
 CC protein isoforms (BPIs). Also described: (1) an antibody capable of  
 CC immunospecifically binding to one of the BPIs; (2) a pharmaceutical  
 CC compositions comprising: (a) a BPI, or a nucleic acid encoding a BPI, and  
 CC a carrier; or (b) the antibody of (1), or a fragment or derivative of the  
 CC antibody, and a carrier; (3) screening for agents that interact with one  
 CC or more BPIs, BPI fragments, polypeptides related to BPIs, or BPI-fusion  
 CC proteins; (4) screening for or identifying agents that modulate the  
 CC expression or activity of one or more BPIs, a BPI fragment, a BPI-related  
 CC polypeptide, or BPI-fusion proteins; and (5) treating or preventing  
 CC breast cancer. BPIs have cytostatic activity and can be used in gene  
 CC therapy. Methods and Kits comprising antibodies or the BPIs from the  
 CC present invention can be used for screening, diagnosing or determining  
 CC the stage or severity of breast cancer, identifying a subject at risk of  
 CC developing breast cancer, or monitoring the effect of therapy  
 CC administered to a subject with breast cancer. The antibodies, BPIs,  
 CC nucleic acids encoding the BPIs, or an agent that modulates the activity  
 CC of one or more BPIs are useful for treating or preventing breast cancer.  
 CC ABP57104 to ABP57250 represent breast cancer associated tryptic digest  
 CC peptides, which are used in the exemplification of the present invention.  
 XX  
 SQ Sequence 16 AA;  
 XX  
 QY  
 Db 2 SPEQGETVLDGNLIR 17  
 1 SPEQGETVLDGNLIR 16  
 Query Match 86.8%; Score 79; DB 24; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-07;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 RESULT 9  
 ABB09706  
 ID ABB09706 standard; Protein; 932 AA.  
 XX  
 AC ABB09706;  
 XX  
 DT 11-JUN-2002 (first entry)  
 XX  
 DE Sequence of H4P heavy chain of inter-alpha-inhibitor protein.  
 XX  
 KM MAY-1; blood transfusion; immunosuppressive; mixed lymphocyte reaction;  
 KM MMR; autoimmune disease; allergic disease; organ rejection;  
 KM organ transplantation; rheumatism; psoriasis; bronchial asthma;  
 KM allergic rhinitis; allergic dermatitis; pollinosis; H4P heavy chain;  
 KM inter-alpha-inhibitor protein.  
 XX  
 OS Rattus sp.  
 XX

PN W0200212495-A1.  
 XX  
 PD 14-FEB-2002.  
 XX  
 PF 01-AUG-2001; 2001WO-JP06620.  
 XX  
 PR 09-AUG-2000; 2000JP-0241169.  
 XX  
 PA (MARU-) MARUHO KK.  
 XX  
 PI Uchida H, Tanaka H, Kitch Y, Fujimura A, Kobayashi E;  
 XX WPI; 2002-217191/27.  
 DR N-PSDB; ABL41969.  
 XX  
 PT Rat or human protein MAY-1 induced by homogeneous blood transfusion and  
 PT exhibiting immunosuppressive activity in a homogeneous mixed lymphocyte  
 PT reaction, for providing drug compositions to treat e.g. autoimmune  
 PT diseases -  
 XX  
 PS Disclosure; Page 43-47; 85pp; Japanese.  
 XX  
 CC The present sequence encodes the rat H4P heavy chain of  
 CC inter-alpha-inhibitor protein. The specification describes MAY-1  
 CC protein, which is induced by homogeneous blood transfusion. MAY-1  
 CC exhibits an immunosuppressive activity in a homogeneous mixed  
 CC lymphocyte reaction (MLR). The MAY-1 protein can formulated into drug  
 CC compositions for the prevention or treatment of autoimmune diseases,  
 CC allergic diseases, or rejection reaction during organ transplantation,  
 CC e.g. rheumatism, psoriasis, bronchial asthma, allergic rhinitis,  
 CC allergic dermatitis and pollinosis.  
 XX  
 SQ Sequence 932 AA;  
 XX  
 QY  
 Db 1 KSPEQGETVLDGNLIR 18  
 223 KSQNEQDTVDGDFVRY 240  
 Query Match 62.6%; Score 57; DB 23; Length 932;  
 Best Local Similarity 55.6%; Pred. No. 0.52;  
 Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 XX  
 RESULT 10  
 ABB09707  
 ID ABB09707 standard; Protein; 933 AA.  
 XX  
 AC ABB09707;  
 XX  
 DT 11-JUN-2002 (first entry)  
 XX  
 DE Sequence of H4P heavy chain of inter-alpha-inhibitor protein.  
 XX  
 KM MAY-1; blood transfusion; immunosuppressive; mixed lymphocyte reaction;  
 KM MMR; autoimmune disease; allergic disease; organ rejection;  
 KM organ transplantation; rheumatism; psoriasis; bronchial asthma;  
 KM allergic rhinitis; allergic dermatitis; pollinosis; H4P heavy chain;  
 KM inter-alpha-inhibitor protein.  
 XX  
 OS Rattus sp.  
 XX  
 FH Key Location/Qualifiers  
 FH Misc-difference 96 /note= "Ieu encoded by AAA"  
 FT Misc-difference 106 /note= "Tyr encoded by ACT"  
 FT  
 FT  
 XX W0200212495-A1.  
 XX  
 PD 14-FEB-2002.  
 XX  
 PF 01-AUG-2001; 2001WO-JP06620.  
 XX

PR 09-AUG-2000; 2000JP-0241169.  
XX (MARU-) MARUHO KK.  
XX Uchida H, Tanaka H, Kitoh Y, Fujimura A, Kobayashi E;  
XX WPI; 2002-217191/27.  
DR N-PSDB; ABL41970.  
XX  
XX Rat or human protein MAY-1 induced by homogeneous blood transfusion and  
PT exhibiting immunosuppressive activity in a homogeneous mixed lymphocyte  
PT reaction, for providing drug compositions to treat e.g. autoimmune  
PT diseases -  
XX  
XX Claim 3; Page 49-53; 85pp; Japanese.  
XX  
XX The present sequence encodes the rat H4p heavy chain of  
CC inter-alpha-inhibitor protein. The specification describes MAY-1  
CC protein, which is induced by homogeneous blood transfusion. MAY-1  
CC exhibits an immunosuppressive activity in a homogeneous mixed  
CC lymphocyte reaction (MLR). The MAY-1 protein can formulated into drug  
CC compositions for the prevention or treatment of autoimmune diseases,  
CC allergic diseases, or rejection reaction during organ transplantation,  
CC e.g. rheumatism, psoriasis, bronchial asthma, allergic rhinitis,  
CC allergic dermatitis and pollinosis.  
XX  
XX Sequence 933 AA;  
SQ  
  
Query Match 62.6%; Score 57; DB 23; Length 933;  
Best Local Similarity 55.6%; Pred. No. 0.53;  
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 KSPGQETVLDDGNLIIRY 18  
Db |||:|||||:|  
224 KSGNEQDTVLDDGFTVRY 241  
  
RESULT 11  
ABP25414 ID ABP25414 standard; Protein; 179 AA.  
XX  
XX ABP25414;  
XX  
XX 02-JUL-2002 (first entry)  
XX  
XX Streptococcus polypeptide SEQ ID NO 4.  
DE  
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.  
XX  
XX Streptococcus agalactiae.  
OS  
XX  
XX WO200234771-A2.  
PN  
XX  
XX 02-MAY-2002.  
PD  
XX  
XX 29-OCT-2001; 2001WO-GB04789.  
PF  
XX  
XX 27-OCT-2000; 2000GB-0026333.  
PR 24-NOV-2000; 2000GB-0028727.  
ER 07-MAR-2001; 2001GB-0005640.  
PR  
XX  
XX (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;  
PI Tettelin H;  
XX  
XX WPI; 2002-352536/38.  
DR N-PSDB; ABN66045.  
XX  
XX New Streptococcus protein for the treatment or prevention of infection  
PT

PT or disease caused by Streptococcus bacteria, such as meningitis, and  
PT for detecting a compound that binds to the protein -  
XX  
XX Claim 1; Page 3154; 4525pp; English.  
XX  
XX The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins.  
XX  
XX Sequence 179 AA;  
SQ  
  
Query Match 50.5%; Score 46; DB 23; Length 179;  
Best Local Similarity 52.9%; Pred. No. 6;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;  
  
QY 2 SPBGQETVLDDGNLIIRY 18  
Db |||:|||||:|  
147 SPENQEKVADNYVVSRY 163  
  
RESULT 12  
ABP29767 ID ABP29767 standard; Protein; 179 AA.  
XX  
XX ABP29767;  
XX  
XX 02-JUL-2002 (first entry)  
XX  
XX Streptococcus polypeptide SEQ ID NO 8710.  
DE  
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.  
XX  
XX Streptococcus agalactiae.  
OS  
XX  
XX WO200234771-A2.  
PN  
XX  
XX 02-MAY-2002.  
PD  
XX  
XX 29-OCT-2001; 2001WO-GB04789.  
PF  
XX  
XX 27-OCT-2000; 2000GB-0026333.  
PR 24-NOV-2000; 2000GB-0028727.  
ER 07-MAR-2001; 2001GB-0005640.  
PR  
XX  
XX (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;  
PI Tettelin H;  
XX  
XX WPI; 2002-352536/38.  
DR N-PSDB; ABN70398.  
XX  
XX New Streptococcus protein for the treatment or prevention of infection  
PT or disease caused by Streptococcus bacteria, such as meningitis, and  
PT for detecting a compound that binds to the protein -  
XX

PS Claim 1; Page 3986; 4525bp; English.

XX

CC The invention relates to a protein (ABP25413-ABP30895) from group B

CC streptococcus pyogenes (Streptococcus agalactiae) or group A streptococcus/GAS

CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in

CC the specification. The proteins have antibacterial and antiinflammatory

CC activity. (1), nucleic acids encoding (1), ABN6044-ABN71526 and

CC antibodies that bind (1) are used in the manufacture of medicaments for

CC the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

CC Nucleic acids encoding (1) are used to detect Streptococcus in a

CC biological sample. (1) is used to determine whether a compound binds to

CC (1). A composition comprising (1) or a nucleic acid encoding (1), may be

CC used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic

CC acid encoding (1) may be used to recombinantly produce (1) and may be

CC used in gene therapy. Antibodies to (1) are used for affinity

CC chromatography, immunoassays, and distinguishing/identifying

CC Streptococcus proteins.

CC

XX

SQ Sequence 179 AA;

Query Match 50.5%; Score 46; DB 23; Length 179;

Best Local Similarity 52.9%; Pred. No. 6;

Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 2 SPEQGETVLDGNLIRY 18

Db 147 SPENQEKVADNVVSRV 163

RESULT 13

ABB92562

ID ABB92562 standard; Protein: 452 AA.

XX

AC ABB92562;

XX

DT 31-MAY-2002 (first entry)

XX

DE Herbicidally active polypeptide SEQ ID NO 1773.

XX

KM Herbicidal; plant; agriculture; herbicide.

XX

OS Arabidopsis thaliana.

XX

PN WO200210210-A2.

XX

PD 07-FEB-2002.

XX

PF 28-AUG-2001; 2001WO-EP09892.

XX

PR 28-AUG-2001; 2001WO-EP09892.

XX

PA (FARB ) BAYER AG.

XX

PI Tietjen K, Weidler M;

XX

DR WPI; 2002-269010/31.

XX

PT Identifying plant target proteins for herbicidally active compounds,

PT comprising aligning and comparing nucleic acid or amino acid sequences

PT from plant with nucleic acid or amino acid sequences from non-plant

PT organisms -

XX

PS Claim 5; SEQ ID NO 1773; 261pp + Sequence Listing; English.

XX

CC The invention relates to identifying target proteins

CC (ABB90790-ABB94016) for herbicidally active compounds, comprising

CC aligning and comparing nucleic acid or amino acid sequences from plant

CC with nucleic acid or amino acid sequences from non-plant organisms using

CC suitable search parameters, where plant sequences having an E-value

CC greater by a factor of 3 than the E-value of most similar non-plant

CC sequences are selected. The polypeptides or nucleic acids encoding them

CC are useful for identifying modulators. The identified modulators are

CC useful as herbicides.

XX

SQ Sequence 452 AA;

Query Match 49.5%; Score 45; DB 23; Length 452;

Best Local Similarity 50.0%; Pred. No. 28;

Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 KSPEQGETVLDGNLIRY 18

Db 160 KDEPKQKVLSGLHPRV 177

RESULT 14

AA41227

ID AA41227 standard; Protein: 409 AA.

XX

AC AA41227;

XX

DT 25-MAR-2003 (updated)

DT 15-MAR-1994 (first entry)

XX

DE 910 SLG protein.

XX

KM Self-incompatibility; Brassica napus rapifera;

KM Brassica campestris; vector; S-locus; amplification;

KM 910 allele; A14 allele; SLG; self linked glycoproteins;

KM plant cells; plant protoplasts.

XX

OS Brassica campestris.

XX

XX

FH Key Location/Qualifiers

FT Peptide 1..31

FT /label= sig\_peptide

FT Protein 32..409

FT /label= mat\_protein

FT Modified-site 48

FT /label= N-glycosylation\_site

FT Modified-site 116

FT /label= N-glycosylation\_site

FT Modified-site 123

FT /label= N-glycosylation\_site

FT Modified-site 264

FT /label= N-glycosylation\_site

FT Modified-site 393

FT /label= N-glycosylation\_site

XX

PN WO9318149-A1.

XX

PD 16-SEP-1993.

XX

XX

PF 29-JUN-1992; 92WO-US04530.

XX

PR 03-MAR-1992; 92US-0847564.

XX

PA (PION-) PIONEER HI-BRED INT INC.

XX

PI Baerzcymski C, Fallis L, Goring DR, Rothstein SJ;

XX

DR WPI; 1993-303463/38.

XX

DR N-P5DB; AAQ48746.

XX

PT Isolated cDNA sequences corresp. to self-incompatibility alleles

PT in Brassica campestris and B napus ssp. rapifera - are useful for

PT transferring self-incompatibility phenotype to plant cells and

PT protoplasts

XX

PS Claim 2; Fig 9; 64pp; English.

XX

CC The cDNAs encoding self incompatibility sites of B. napus and

CC B. campestris (AAQ48745-46) correspond to genes encoding S-linked

CC glycoproteins which are associated with the expression of

CC sporophytic self-incompatibility. The cDNAs may be used to  
CC transform self-compatible plants, plant cells and plant  
CC protoplasts, with the aim of introducing the self-incompatibility  
CC phenotype.  
CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 409 AA;

Query Match 48.4%; Score 44; DB 14; Length 409;  
Best Local Similarity 44.4%; Pred. No. 38;  
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 KSPEDQETVLGNTLIRY 18  
Db 130 RSPVVAELLANGNFVIRY 147

RESULT 15

AY85180  
ID AY85180 standard; Protein; 650 AA.

AC AY85180;

DT 29-JUN-2000 (first entry)

DE Cellulose synthase subunit amino acid sequence.

KW Cellulose synthase; cellulose production; increase yield.

OS Vigna angularis.

FT Key Location/Qualifiers

FT Misc-difference 379 /label= Unknown

FT /note= "Encoded by GTN"

PN JP2000060568-A.

PD 29-FEB-2000.

PF 26-AUG-1998; 98JP-0239998.

PR 26-AUG-1998; 98JP-0239998.

PA (MIZU/) MIZUNO K.

PA (OJIP ) OJI PAPER CO.

DR WPI; 2000-342371/30.

DR N-PSDB; AAA10595.

PT A gene encoding a cellulose synthetic equipment - for the improvement

PT in the amount of cellulose synthesised in a plant body

PS Claim 1; Page 11-14; 32pp; Japanese.

XX This sequence represents the amino acid sequence of a subunit of the  
CC cellulose synthase complex of *Vigna angularis*. The invention relates to  
CC subunits of cellulose synthetic equipment, that can be used to increase  
CC the amount of cellulose synthesised by a plant. The proteins and genes  
CC encoding them can also be used to improve the properties of the cellulose  
CC being produced by a plant.

XX Sequence 650 AA;

Query Match 47.3%; Score 43; DB 21; Length 650;

Best Local Similarity 69.2%; Pred. No. 1e+02; Mismatches 3; Indels 0; Gaps 0;

QY 2 SPEQOETVLGNTL 14

Db 563 SPEQORTVLGNTL 575

RESULT 16

ABP27072  
ID ABP27072 standard; Protein; 858 AA.

AC ABP27072;

DT 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 3320.

KW Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;

KW group A streptococcus; Streptococcus pyogenes; antibacterial;

KW antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus agalactiae.

PN WO200234771-A2.

PD 02-MAY-2002.

PF 29-OCT-2001; 2001WO-GB04789.

PR 27-OCT-2000; 2000GB-0026333.

PR 24-NOV-2000; 2000GB-0028727.

PR 07-MAR-2001; 2001GB-0005640.

PA (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Frazer C;

PI Tettelin H;

DR WPI; 2002-352536/38.

DR N-PSDB; ABN67703.

PT New Streptococcus protein for the treatment or prevention of infection

PT or disease caused by Streptococcus bacteria, such as meningitis, and

PT for detecting a compound that binds to the protein -

PS Claim 1; Page 3486; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly *S. agalactiae* and *S. pyogenes*.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins.

XX Sequence 858 AA;

Query Match 47.3%; Score 43; DB 23; Length 858;

Best Local Similarity 50.0%; Pred. No. 1.4e+02; Mismatches 3; Indels 0; Gaps 0;

QY 2 SPEQOETVLGNTL 15

Db 400 APEAQATITTEGNTL 413

RESULT 17

AB29100  
ID AB29100 standard; Protein; 946 AA.

XX AAB29100;  
AC  
XX 07-FEB-2001 (first entry)  
DT  
XX  
XX Human inter-alpha-tryptase inhibitor heavy-chain II precursor.  
DE  
XX Human; inter-alpha-tryptase inhibitor heavy-chain II precursor;  
KW cell migration; inflammation; autoimmune disease; fugalactic agent.  
XX  
OS Homo sapiens.  
XX W0200059941-A1.  
XX 12-OCT-2000.  
XX  
XX 07-APR-2000; 2000MO-US09678.  
XX 08-APR-1999; 99US-0128272.  
XX 03-DEC-1999; 99US-0168952.  
XX (GENO ) GEN HOSPITAL CORP.  
XX  
XX Poznański MC, Luster AD, Scadden DT;  
XX WPI; 2000-672621/65.  
XX  
XX Isolates from thymic stromal-cell, HepG2-cell or Kaposi's sarcoma  
PT Hepes virus infected-cell useful for treating tumour, infertility and  
PT premature labour, repel immune cells -  
XX  
XX Claim 11; Page 83-85; 87pp; English.  
XX  
XX The present invention is related to the modulation of movement of  
CC eukaryotic cells. This involves the use of fugalactic agents, such as the  
CC one shown here. These can be used in the treatment of inflammation and  
CC autoimmune diseases, including rheumatoid arthritis, uveitis, diabetes,  
CC haemolytic anaemias, rheumatic fever, Crohn's disease, Guillain-Barre  
CC syndrome, psoriasis, thyroiditis, Graves' disease, myasthenia gravis,  
CC glomerulonephritis, autoimmune hepatitis and systemic lupus  
CC erythematosus. They can also be used to treat cancer, infertility and  
CC premature labour.  
XX  
XX Sequence 946 AA;  
SQ  
Query Match 47.3%; Score 43; DB 21; Length 946;  
Best Local Similarity 43.8%; Pred. No. 1.6e+02;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
OY 3 PEOQETVLDGNTLIRY 18  
DB 262 PSCHETAVDGLVLY 277

RESULT 18  
ABP70358  
ID ABP70358 standard; peptide: 28 AA.  
XX  
XX ABP70358;  
XX  
XX 07-APR-2003 (first entry)  
XX  
XX Tryptic peptide of heat shock protein 60 of Chlamydia pneumoniae.  
DE  
XX Microorganism; mass spectroscopy; biomarker; heat shock protein 60;  
KW Hep60.  
XX  
XX Chlamydia pneumoniae.  
OS  
XX W0200295416-A2.  
XX  
XX 28-NOV-2002.  
XX

PF 22-MAY-2002; 2002MO-GB02208.  
XX  
XX 22-MAY-2001; 2001GB-0012428.  
XX  
XX (MINA ) UK SEC FOR DEFENCE.  
XX  
XX Tiltball RW, Despeyroux D;  
XX  
XX WPI; 2003-129474/12.  
XX  
XX New biomarker comprising species homologues derived from the majority  
PT of species in at least two groups of microorganisms that are  
PT structurally similar, useful for identifying microorganisms -  
XX  
XX Example 2; Page 17; 35pp; English.  
XX  
XX The specification describes a method of rapidly identifying unknown  
CC microorganisms by means of mass spectroscopy of biomarkers. The  
CC biomarkers comprise species homologues derived from the majority of  
CC species in at least two groups of microorganisms that are structurally  
CC similar. The structural similarity allows isolation of the biomarkers  
CC from different species of microorganism. Each biomarker has an unique  
CC molecular mass. The biomarker is useful for identifying microorganisms.  
CC Peptides ABP70357-91 represent a tryptic peptide map of the heat shock  
CC protein 60 (Hsp60) of Chlamydia pneumoniae. Hsp60 is used as a biomarker  
CC to demonstrate the method of the invention.  
XX  
XX Sequence 28 AA;  
SQ  
Query Match 46.2%; Score 42; DB 24; Length 28;  
Best Local Similarity 52.9%; Pred. No. 3;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
OY 2 SPEQETVLDGNTLIRY 18  
DB 10 NPETOECVLDALIRY 26

RESULT 19  
AAM10977  
ID AAM10977 standard; Protein; 419 AA.  
XX  
XX AAM10977;  
XX  
XX 21-MAY-1997 (first entry)  
XX  
XX Dihydrofolate reductase-Chlamydia pneumoniae antigen fusion protein.  
DE  
XX DHPR; dihydrofolate reductase; Chlamydia pneumoniae; pneumonia;  
KW antibody production; diagnosis; fusion protein.  
XX  
XX Chlamydia pneumoniae (chimeric).  
XX  
XX  
XX Key Location/Qualifiers  
FH Region 1..160  
FT /note= "dihydrofolate reductase region"  
FT Region 161..170  
FT /note= "peptide linker"  
FT Misc-difference 171..407  
FT /note= "Chlamydia pneumoniae antigen region"  
XX  
XX JF08294391-A.  
XX  
XX 12-NOV-1996.  
XX  
XX 28-APR-1995; 95JP-0106007.  
XX  
XX 28-APR-1995; 95JP-0106007.  
XX  
XX (HITB ) HITACHI CHEM CO LTD.  
XX  
XX WPI; 1997-036901/04.  
XX

PT Fusion protein comprising dihydrofolate reductase and Chlamydia  
 PT pneumoniae antigen - useful in prodn. of C. pneumoniae antibodies  
 PT for diagnosis of infection  
 XX  
 PS Claim 5; Page 14-15; 17pp; Japanese.  
 XX  
 CC AAW10977 is a dihydrofolate reductase (DHFR)-Chlamydia pneumoniae  
 CC antigen fusion protein. Fusion proteins that may be made consist of at  
 CC least 5 contiguous amino acids of the Chlamydia pneumoniae antigen  
 CC linked to the N-terminus of a DHFR protein (see AAW10974). Fusion  
 CC proteins produced are useful for the production of anti-C. pneumoniae  
 CC antibodies which are useful in the diagnosis and treatment of infectious  
 CC diseases caused by C. pneumoniae.  
 CC  
 SQ Sequence 419 AA;  
 Query Match 46.2%; Score 42; DB 18; Length 419;  
 Best Local Similarity 52.9%; Pred. No. 88;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 QY 2 SPEQETVLDGNLIIRY 18  
 :||| |||: |||:  
 Db 175 NPETQECVLEDPALILY 191  
 RESULT 20  
 AAW1865  
 ID AAW1865 standard; Protein; 419 AA.  
 XX  
 AC AAW1865;  
 XX  
 DT 17-APR-1997 (first entry)  
 XX  
 DE DHFR/Polypeptide B fusion protein #2.  
 XX  
 KW Dihydrofolate reductase; DHFR; C. pneumoniae; detection; antibody;  
 KW fusion protein; antigen; diagnosis.  
 XX  
 OS Chlamydia pneumoniae.  
 XX  
 FH Key Location/Qualifiers  
 FT Protein 1..160  
 FT /label= DHFR  
 FT 172..407  
 FT /label= Polypeptide\_B\_residues\_203-439  
 XX  
 PM JP08304403-A.  
 XX  
 PD 22-NOV-1996.  
 XX  
 PF 28-APR-1995; 95JP-0106012.  
 XX  
 PR 28-APR-1995; 95JP-0106012.  
 XX  
 PA (HITB ) HITACHI CHEM CO LTD.  
 XX  
 DR WPI; 1997-056177/06.  
 XX  
 PT Detection and determination of anti-Chlamydia pneumoniae antibody -  
 PT using an antigenic polypeptide fused to the Chlamydia  
 PT dihydrofolate reductase as the antigen  
 XX  
 PS Claim 5; Page 15-16; 17pp; Japanese.  
 XX  
 CC This sequence represents a fusion between C. pneumoniae polypeptide  
 CC B and dihydrofolate reductase (DHFR). This protein was used in the  
 CC method of the invention for the detection and determination of  
 CC anti-Chlamydia pneumoniae antibody. DHFR and polypeptide B were  
 CC combined directly or through an amino acid sequence to give a fusion  
 CC protein to act as an antigen. The fusion protein may be used in  
 CC a reagent which has a high reliability and which gives an exact  
 CC diagnosis.  
 CC

SQ Sequence 419 AA;  
 Query Match 46.2%; Score 42; DB 18; Length 419;  
 Best Local Similarity 52.9%; Pred. No. 88;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 QY 2 SPEQETVLDGNLIIRY 18  
 :||| |||: |||:  
 Db 175 NPETQECVLEDPALILY 191  
 RESULT 21  
 AAW10975  
 ID AAW10975 standard; Protein; 544 AA.  
 XX  
 AC AAW10975;  
 XX  
 DT 21-MAY-1997 (first entry)  
 XX  
 DE Chlamydia pneumoniae antigen used as DHFR-linked fusion protein.  
 XX  
 KW DHFR; dihydrofolate reductase; Chlamydia pneumoniae; pneumonia;  
 KW antibody production; diagnosis; fusion protein.  
 XX  
 OS Chlamydia pneumoniae.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 25  
 FT /note= "given as G1e in three letter amino  
 FT acid code in the specification"  
 XX  
 PM JP08294391-A.  
 XX  
 PD 12-NOV-1996.  
 XX  
 PF 28-APR-1995; 95JP-0106007.  
 XX  
 PR 28-APR-1995; 95JP-0106007.  
 XX  
 PA (HITB ) HITACHI CHEM CO LTD.  
 XX  
 DR WPI; 1997-036901/04.  
 XX  
 PT Fusion protein comprising dihydrofolate reductase and Chlamydia  
 PT pneumoniae antigen - useful in prodn. of C. pneumoniae antibodies  
 PT for diagnosis of infection  
 XX  
 PS Claim 1; Page 11-12; 17pp; Japanese.  
 XX  
 CC AAW10974 encodes a 544 residue Chlamydia pneumoniae antigen, at least  
 CC 5 contiguous amino acids of which are fused to a dihydrofolate  
 CC reductase (DHFR) enzyme. Fusion proteins produced are useful for the  
 CC production of anti-C. pneumoniae antibodies which are useful in the  
 CC diagnosis and treatment of infectious diseases caused by C. pneumoniae.  
 CC  
 SQ Sequence 544 AA;  
 Query Match 46.2%; Score 42; DB 18; Length 544;  
 Best Local Similarity 52.9%; Pred. No. 1.2e+02;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 QY 2 SPEQETVLDGNLIIRY 18  
 :||| |||: |||:  
 Db 207 NPETQECVLEDPALILY 223  
 RESULT 22  
 AAW1863  
 ID AAW1863 standard; Protein; 544 AA.  
 XX  
 AC AAW1863;  
 XX  
 DT 17-APR-1997 (first entry)

XX Polypeptide B.  
 DE Dihydrofolate reductase; DHFR; C. pneumoniae; detection; antibody;  
 XX fusion protein; antigen; diagnosis.  
 XX Chlamydia pneumoniae.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 25 /note= "Given in the specification as G1e"  
 XX  
 PN JF08304403-A.  
 XX  
 PD 22-NOV-1996.  
 XX  
 PF 28-APR-1995; 95JP-0106012.  
 XX  
 PR 28-APR-1995; 95JP-0106012.  
 XX  
 PA (HITB ) HITACHI CHEM CO LTD.  
 XX  
 DR WPI; 1997-056177/06.  
 XX  
 PT Detection and determination of anti-Chlamydia pneumoniae antibody -  
 PT using an antigenic polypeptide fused to the Chlamydia  
 PT dihydrofolate reductase as the antigen  
 XX  
 PS Claim 1; Page 11-13; 17pp; Japanese.  
 XX  
 CC This sequence represents C. pneumoniae polypeptide B. Polypeptide B  
 CC was linked to dihydrofolate reductase (DHFR) in the method of the  
 CC invention for the detection and determination of anti-Chlamydia  
 CC pneumoniae antibody. DHFR and polypeptide B were combined directly  
 CC or through an amino acid sequence to give a fusion protein as in  
 CC AAW1864-65, to act as an antigen. The fusion protein may be used in  
 CC a reagent which has a high reliability and which gives an exact  
 CC diagnosis.  
 CC  
 SQ Sequence 544 AA;  
 XX  
 QY Query Match 46.2%; Score 42; DB 18; Length 544;  
 Best Local Similarity 52.9%; Pred. No. 1.2e+02;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 XX  
 Db 2 SPEQGETVLDGNLIRY 18  
 :||| |||: |||:  
 207 NPEQGEVLDPALILY 223  
 XX  
 RESULT 23  
 AAY34735  
 ID AAY34735 standard; Protein; 544 AA.  
 XX  
 AC AAY34735;  
 XX  
 DT 13-SEP-1999 (first entry)  
 XX  
 DE Amino acid sequence of a Chlamydia pneumoniae protein.  
 XX  
 KM Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KM sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KM vaccine; neutralising epitope.  
 XX  
 OS Chlamydia pneumoniae.  
 XX  
 PN WO9927105-A2.  
 XX  
 PD 03-JUN-1999.  
 XX  
 PF 20-NOV-1998; 98WO-IB01890.  
 XX  
 PR 04-NOV-1998; 98US-0107078.  
 XX

PR 21-NOV-1997; 97FR-0014673.  
 XX  
 PA (GERT ) GENSET.  
 XX  
 PI Griffiths R;  
 XX  
 DR WPI; 1999-357842/30.  
 XX  
 PT Genome sequence of Chlamydia pneumoniae  
 XX  
 PS Page 726-727; Disclosure; 1912pp; English.  
 XX  
 CC AAY34584-Y35879 represent the proteins encoded by all the open reading  
 CC frames in the complete genome (see AAY31990) of Chlamydia pneumoniae.  
 CC C. pneumoniae causes respiratory disease such as pneumonia and  
 CC bronchitis and is thought to be a contributing factor in heart  
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in  
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
 CC nucleotide sequences can also be used as immunogenic compositions,  
 CC especially where the vector directs the expression of a neutralising  
 CC epitope of C. pneumoniae.  
 CC  
 SQ Sequence 544 AA;  
 XX  
 QY Query Match 46.2%; Score 42; DB 20; Length 544;  
 Best Local Similarity 52.9%; Pred. No. 1.2e+02;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 XX  
 Db 2 SPEQGETVLDGNLIRY 18  
 :||| |||: |||:  
 207 NPEQGEVLDPALILY 223  
 XX  
 RESULT 24  
 AAE11757  
 ID AAE11757 standard; Protein; 544 AA.  
 XX  
 AC AAE11757;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Chlamydia pneumoniae heat shock protein 60 (HSP60).  
 XX  
 KM Heat shock protein 60; HSP60; antiarteriosclerotic; antiinflammatory;  
 KM antiallergic; immunomodulator; dermatological; immune response;  
 KM vasoactive; immunostimulant; therapy; vascular disorder; immune response;  
 KM atherosclerosis; allergic angitis; Behcet's syndrome; granulomatosis;  
 KM Churg-Straus disease; Cogan's syndrome; graft-versus-host disease; GVHD;  
 KM Henoch-Schönlein purpura; leucocytoclastic vasculitis; Kawasaki disease;  
 KM polyarteritis nodosa; PAN; Takayasu's arteritis; temporal arteritis;  
 KM thromboangiitis obliterans; Wegener's disease; transplant rejection;  
 KM microscopic polyangitis.  
 XX  
 OS Chlamydia pneumoniae.  
 XX  
 PN WO200168124-A2.  
 XX  
 PD 20-SEP-2001.  
 XX  
 PF 15-MAR-2001; 2001WO-US08351.  
 XX  
 PR 15-MAR-2000; 2000US-189855P.  
 XX  
 PA (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.  
 XX  
 PI Weiner HL, Maron R, Libby P;  
 XX  
 DR WPI; 2001-611383/70.  
 XX  
 PT Treating a vascular disorder, involves administering a composition

PT comprising heat shock protein, its fragment or analog, by mucosal  
PT surface, pulmonary tract, oral or enteral route, or by inhalation  
XX  
PS Claim 10; Page 12-13; 49pp; English.  
XX  
CC The patent discloses methods for treating vascular disorders in  
CC mammals. The method involves administering a composition comprising  
CC at least one agent selected from heat shock protein (HSP), its fragment  
CC or analogue, through mucosal surface, pulmonary tract, oral or enteral  
CC route or by inhalation. Compositions comprising HSP are useful for  
CC treating and suppressing a vascular disorder, including cell-mediated  
CC immune response, an antibody-mediated immune response, cell-mediated  
CC inflammatory disorder, atherosclerosis, allergic angitis, Behcet's  
CC syndrome, granulomatosis (Churg-Straus disease), Cogan's syndrome,  
CC graft-versus-host disease (GVHD), Henoch-Schönlein purpura, Kawasaki  
CC disease, leucocytoclastic vasculitis, polyarteritis nodosa (PAN),  
CC microscopic polyangitis, polyangitis overlap syndrome, Takayasu's  
CC arteritis, temporal arteritis, transplant rejection, Wegener's  
CC granulomatosis and thromboangiitis obliterans (Buerger's disease).  
CC They are useful for reducing the level of proinflammatory Th1 cytokines  
CC and also for increasing the level of antiinflammatory Th2 cytokines.  
CC The present sequence is heat shock protein 60 (HSP60) from  
CC Chlamydia pneumoniae.

XX Sequence 544 AA;

Query Match 46.2%; Score 42; DB 22; Length 544;  
Best Local Similarity 52.9%; Pred. No. 1.2e+02;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Oy 2 SPEOQETVLDGNLIIRY 18

Db 207 NPTQECVLEDAIILY 223

RESULT 25

ABR94272 ID ABR94272 standard; Protein; 544 AA.

XX ABR94272;

XX 05-JUN-2002 (first entry)

DE Chlamydia pneumoniae protein sequence SEQ ID NO:400.

XX Chlamydial infection; Chlamydia; vaccine; detection; diagnosis;

KW antigen; antibacterial; immunostimulant; immune response;

KW Chlamydia-specific T-cell response.

XX Chlamydia pneumoniae.

XX WO200208267-A2.

PD 31-JAN-2002.

PF 20-JUN-2001; 2001WO-US23121.

PR 20-JUN-2000; 2000US-0620412.

PR 23-APR-2001; 2001US-0841132.

PA (CORI-) CORIXA CORP.

PI Fling SP, Skeiky YAW, Probst P, Bhatia A;

DR WPI; 2002-179901/23.

XX Novel compositions comprising Chlamydia Cap1 protein and its use in the

PT treatment of Chlamydia infection -

XX Disclosure; Page 363-364; 537pp; English.  
XX The present invention describes compositions comprising a Chlamydia Cap1  
CC protein and methods for the diagnosis and therapy of Chlamydia infection.

CC Chlamydia DNA and protein sequences from the present invention can have  
CC antibacterial and immunostimulant activities, and can be used in  
CC vaccines. Compounds from the present invention can be used for eliciting  
CC an immune response, specifically stimulating a Chlamydia-specific T-cell  
CC response or inhibiting the development of a Chlamydia infection in an  
CC animal. Methods from the present invention can be used: for detecting the  
CC presence of Chlamydia in a patient; to stimulate and/or expand T cells  
CC specific for a Chlamydia protein; and for treatment of a Chlamydia  
CC infection. ABL92394 to ABL92709 and ABR94096 to ABR94374 represent  
CC sequences used in the exemplification of the present invention.

XX Sequence 544 AA;

Query Match 46.2%; Score 42; DB 23; Length 544;  
Best Local Similarity 52.9%; Pred. No. 1.2e+02;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Oy 2 SPEOQETVLDGNLIIRY 18

Db 207 NPTQECVLEDAIILY 223

RESULT 26

AAE13000 ID AAE13000 standard; Protein; 652 AA.

XX AAE13000;

XX 28-JAN-2002 (first entry)

DE Corn poly (A) binding protein from clone p0083.c1dcq55r.

XX Corn; polyadenylated RNA-binding protein; transgenic plant; herbicide;

KW eIF; eukaryotic translation initiation factor-4 (eIF-4) gamma;

KW genetic mapping; physical mapping.

XX Zea mays.

XX Key Location/Qualifiers

FT Misc-difference 52 /note= "Encoded by GAA"

FT Misc-difference 67 /note= "Encoded by AGG"

XX US6294658-B1.

PD 25-SEP-2001.

PF 02-JUN-1999; 99US-0347833.

PR 10-JUN-1998; 98US-092415P.

PA (DUPO ) DU PONT DE NEMOURS & CO E I.

PI Famodu LO, Odell JT;

DR WPI; 2001-656175/75.

DR N-PSDB; AAD21313.

XX New polynucleotides encoding poly(A) binding protein or a eukaryotic

PT translation initiation factor-4 gamma, useful for creating transgenic

PT plants where the proteins are present at higher or lower levels than

normal

XX Claim 9; Fig 1-3; 26pp; English.

XX The present invention relates to an isolated polynucleotide encoding a

CC polyadenylated RNA-binding protein or eukaryotic translation initiation

CC factor-4 (eIF-4) gamma. The nucleic acid fragments may be used in PCR

CC protocols to amplify longer nucleic acid fragments encoding homologous

CC genes from DNA or RNA, to create transgenic plants in which the new

CC polypeptides are present at higher or lower levels than normal or in cell

CC types or developmental stages in which they are not normally found. The



CC nucleic acids may also be used as probes for genetic or physical mapping  
CC the genes that they are part of and as makers for traits linked to those  
CC genes. The polypeptides may be used to produce antibodies and as targets  
CC to facilitate design and/or identification of inhibitors of those enzymes  
CC that can be used as herbicides. The present sequence is corn poly (A)  
CC binding protein from clone p0083.cldq55r.  
XX

SQ Sequence 652 AA;

Query Match 46.2%; Score 42; DB 22; Length 652;  
Best Local Similarity 61.5%; Pred. No. 1.5e+02;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 SPEQOETVL DGNL 14  
:|||||:|:|  
Db 565 NPEQORTILGNL 577

RESULT 27

AAE13002 ID AAE13002 standard; Protein; 652 AA.

XX AAE13002;

DT 28-JAN-2002 (first entry)

XX Glycine max poly (A) binding protein.

XX Polyadenylated RNA-binding protein; transgenic plant; herbicide;  
KW eukaryotic translation initiation factor-4 (eif-4) gamma;  
XX genetic mapping; physical mapping; soybean.  
XX

OS Glycine max.

XX Key Location/Qualifiers

FT Misc-difference 160 /note= "Encoded by NCTC"

FT Misc-difference 164 /note= "Encoded by GCCC"

FT Misc-difference 167 /note= "Encoded by CAGA"

FT Misc-difference 174..428 /note= "Encoded by GC"

XX US6294658-B1.

XX 25-SEP-2001.

XX 02-JUL-1999; 99US-0347833.

XX 10-JUL-1998; 98US-092415P.

XX (DUPO ) DU PONT DE NEMOURS & CO E I.

XX Famodu LO, . Odel1 JT;

XX WPI; 2001-656175/75.

XX N-PSDB; AAD21315.

XX New polynucleotides encoding poly(A) binding protein or a eukaryotic  
PT translation initiation factor-4 gamma, useful for creating transgenic  
PT plants where the proteins are present at higher or lower levels than  
PT normal

XX Claim 9, Fig 2-3; 26pp; English.

XX The present invention relates to an isolated polynucleotide encoding a  
CC polyadenylated RNA-binding protein or eukaryotic translation initiation  
CC factor-4 (eif-4) gamma. The nucleic acid fragments may be used in PCR  
CC protocols to amplify longer nucleic acid fragments encoding homologous  
CC genes from DNA or RNA, to create transgenic plants in which the new  
CC polypeptides are present at higher or lower levels than normal or in cell  
CC types or developmental stages in which they are not normally found. The

CC nucleic acids may also be used as probes for genetic or physical mapping  
CC the genes that they are part of and as makers for traits linked to those  
CC genes. The polypeptides may be used to produce antibodies and as targets  
CC to facilitate design and/or identification of inhibitors of those enzymes  
CC that can be used as herbicides. The present sequence is soybean poly (A)  
CC binding protein from clone sdp2c.pk003.112.  
XX

SQ Sequence 652 AA;

Query Match 46.2%; Score 42; DB 22; Length 652;  
Best Local Similarity 61.5%; Pred. No. 1.5e+02;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 SPEQOETVL DGNL 14  
:|||||:|:|  
Db 565 NPEQORTILGNL 577

RESULT 28

AAW10976 ID AAW10976 standard; Protein; 704 AA.

XX AAW10976;

DT 21-MAY-1997 (first entry)

XX Dihydrofolate reductase-Chlamydia pneumoniae antigen fusion protein.

XX DHFR, dihydrofolate reductase; Chlamydia pneumoniae; pneumonia;  
KW antibody production; diagnosis; fusion protein.  
XX

OS Chlamydia pneumoniae (chimeric).

XX Key Location/Qualifiers

FT Region 1..160 /note= "dihydrofolate reductase region"

FT Region 161..704 /note= "Chlamydia pneumoniae antigen"

FT Misc-difference 185 /note= "given as G1e in three letter amino  
FT acid code in the specification"

XX JP08294391-A.

XX 12-NOV-1996.

XX 28-APR-1995; 95JP-0106007.

XX 28-APR-1995; 95JP-0106007.

XX (HITB ) HITACHI CHEM CO LTD.

XX WPI; 1997-036901/04.

XX Fusion protein comprising dihydrofolate reductase and Chlamydia  
PT pneumoniae antigen useful in prodn. of C. pneumoniae antibodies  
PT for diagnosis of infection

XX Claim 4; Page 12-14; 17pp; Japanese.

XX AAW10976 is a dihydrofolate reductase (DHFR)-Chlamydia pneumoniae  
CC antigen fusion protein. Fusion proteins that may be made consist of at  
CC least 5 contiguous amino acids of the Chlamydia pneumoniae antigen  
CC linked to the C-terminus of a DHFR protein (see AAW10974). Fusion  
CC proteins produced are useful for the production of anti-C. pneumoniae  
CC antibodies which are useful in the diagnosis and treatment of  
CC infectious diseases caused by C. pneumoniae.

XX SQ Sequence 704 AA;

XX Query Match 46.2%; Score 42; DB 18; Length 704;  
Best Local Similarity 52.9%; Pred. No. 1.7e+02;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;





CC discloses genomic DNA sequences (ABJ16176-ABJ30511), expressed DNA  
CC sequences (ABJ0140-ABJ16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2515 AA;  
Query Match 46.2%; Score 42; DB 22; Length 2515;  
Best Local Similarity 42.9%; Pred. No. 8.1e+02;  
Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;  
QY 1 KSPQOETVLDGNTL 14  
: |||: |||:  
Db 959 QKPOROKSTLDGNI 972  
RESULT 33  
AAG82710  
ID AAG82710 standard; Protein; 346 AA.  
XX  
AC AAG82710;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE S. epidermidis open reading frame protein sequence SEQ ID NO:2514.  
XX  
KW Staphylococcus epidermidis SRI strain; infection; diagnosis;  
XX vaccination; endocarditis.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN W0200134809-A2.  
XX  
PD 17-MAY-2001.  
XX  
PF 09-NOV-2000; 2000WO-US30782.  
XX  
PR 09-NOV-1999; 99US-0164258.  
XX  
PA (GLAXO ) GLAXO GROUP LTD.  
XX  
PI Kimmery WJ;  
XX  
XX WPI; 2001-316495/33.  
DR N-PSDB; AAH53560.  
XX  
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis -  
PS Claim 18; Page 661; 2188pp; English.  
XX  
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG8120, from Staphylococcus epidermidis.  
CC (I) and (II) can have antibacterial activity and therefore can be used  
CC in vaccination. The nucleic acids (I) may be used to produce the  
CC S. epidermidis polypeptides (II) via the production of vectors  
CC containing them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH55091 to  
CC AAH55098 represent oligonucleotide sequences and primers which are used  
CC in the exemplification of the present invention.  
CC N.B. The present invention specifically claims all the polynucleotide  
CC sequences given in the sequence listing of the present specification,  
CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
CC no sequences are present for SEQ ID NO:4455 to 4464.

XX  
SQ Sequence 346 AA;  
Query Match 45.6%; Score 41.5; DB 22; Length 346;  
Best Local Similarity 55.0%; Pred. No. 85;  
Matches 11; Conservative 1; Mismatches 3; Indels 5; Gaps 1;  
QY 2 SPQOETVLD-----GNLII 16  
: |||: |||:  
Db 254 SPQOETVLDVKKTKRLIV 273  
RESULT 34  
ABP39795  
ID ABP39795 standard; Protein; 353 AA.  
XX  
AC ABP39795;  
XX  
DT 24-JUL-2002 (first entry)  
XX  
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4640.  
XX  
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
XX antibacterial; gene therapy.  
XX  
OS Staphylococcus epidermidis.  
XX  
PN US6380370-B1.  
XX  
PD 30-APR-2002.  
XX  
PF 13-AUG-1998; 98US-0134001.  
XX  
PR 14-AUG-1997; 97US-055779P.  
XX  
PR 08-NOV-1997; 97US-064964P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
XX WPI; 2002-381255/41.  
DR N-PSDB; ABN92340.  
XX  
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
PT polypeptide, useful for diagnosing and treating bacterial infections -  
XX  
PS Disclosure; SEQ ID 4640; 267pp; English.  
XX  
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
CC antibacterial activity and can be used in gene therapy. The sequences  
CC can also be used in the diagnosis and treatment of bacterial infections,  
CC particularly S. epidermidis infections. The sequences can be used to  
CC screen for compound able to interfere with the S. epidermidis life  
CC cycle or inhibit S. epidermidis infection.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC USPTO web site.  
XX  
SQ Sequence 353 AA;  
Query Match 45.6%; Score 41.5; DB 23; Length 353;  
Best Local Similarity 55.0%; Pred. No. 87;  
Matches 11; Conservative 1; Mismatches 3; Indels 5; Gaps 1;  
QY 2 SPQOETVLD-----GNLII 16  
: |||: |||:  
Db 261 SPQOETVLDVKKTKRLIV 280  
RESULT 35  
AAG79929

ID	AA679929 standard; Protein: 1254 AA.				FT	Domain	301..319
XX					FT		/label= Coiled coil domain
AC	AA679929;				FT		/note= "Identified by BLAST_DOMO"
XX					FT	Domain	313..510
DT	21-MAY-2003 (first entry)				FT		/label= Coiled coil domain
XX					FT		/note= "Identified by BLAST_DOMO"
DE	CGDD-4, Incyte ID No. 3214362CD1.				FT	Domain	328..1219
XX					FT		/label= Coiled coil domain
KW	Human; cell growth; differentiation; death; CGDD;				FT	Domain	/note= "Identified by BLAST_DOMO"
KW	cell proliferation; cancer; actinic keratosis; arteriosclerosis;				FT		351..469
KW	atherosclerosis; bursters; development; renal tubular acidosis;				FT		/label= Coiled coil domain
KW	anaemia; Cushing's syndrome; Duchenne; Becker; muscular dystrophy;				FT		/note= "Identified by BLAST_DOMO"
KW	neurology; epilepsy; stroke; cerebral neoplasm; Alzheimer's disease;				FT	Domain	353..432
KW	Pick's disease; autoimmune; inflammation; AIDS; allergy; asthma;				FT		/label= Coiled coil domain
KW	thyroiditis; metabolism; obesity; type II diabetes; reproduction;				FT		/note= "Identified by BLAST_DOMO"
KW	fertility; prolactin production; placenta; preeclampsia;				FT	Domain	357..435
KW	choriocarcinoma; abruptio placentae; choriangioma; transgenic animal;				FT		/label= Coiled coil domain
KW	bovine; SMCI protein.				FT		/note= "Identified by BLAST_DOMO"
XX					FT	Domain	362..443
OS	Homo sapiens.				FT		/label= Coiled coil domain
XX					FT		/note= "Identified by BLAST_DOMO"
PH	Key				FT	Domain	365..927
FT					FT		/label= Coiled coil domain
FT					FT		/note= "Identified by BLAST_DOMO"
FT					FT	Domain	365..971
FT	Binding-site				FT		/label= Coiled coil domain
FT					FT		/note= "Identified by BLAST_DOMO"
FT					FT	Domain	380..1027
FT					FT		/label= Coiled coil domain
FT					FT		/note= "Identified by BLAST_DOMO"
FT					FT	Domain	416..1235
FT					FT		/label= Coiled coil domain
FT					FT		/note= "Identified by BLAST_DOMO"
FT					FT	Protein	449..508
FT					FT		/label= Homologous to SMCI chromosome segregation
FT					FT		/note= "Identified by BLAST_PRODOW"
FT					FT	Protein	526..664
FT					FT		/label= Homologous to Coiled coil ATP binding
FT					FT		/note= "Identified by BLAST_PRODOW"
FT					FT	Domain	670..930
FT					FT		/label= Coiled coil domain
FT					FT		/note= "Identified by BLAST_DOMO"
FT					FT	Domain	707..1223
FT					FT		/label= Coiled coil domain
FT					FT		/note= "Identified by BLAST_DOMO"
FT					FT	Domain	751..862
FT					FT		/label= Coiled coil domain
FT					FT		/note= "Identified by BLAST_DOMO"
FT					FT	Domain	904..1031
FT					FT		/label= Coiled coil domain
FT					FT		/note= "Identified by BLAST_DOMO"
FT					FT	Protein	967..1139
FT					FT		/label= Homologous to SMCI chromosome segregation
FT					FT		/note= "Identified by BLAST_PRODOW"
FT					FT	Domain	1021..1235
FT					FT		/label= SWC family C-terminal domain
FT					FT		/note= "Identified by HMMER_PPFAM"
FT					FT	Domain	1051..1219
FT					FT		/label= Coiled coil domain
FT					FT		/note= "Identified by BLAST_DOMO"
XX					PN	W02003008553-A2.	
XX					XX		
PD	30-JAN-2003.				XX		
XX					PF	16-JUL-2002; 2002MO-US22834.	
XX					XX		
PR	17-JUL-2001; 2001US-306064P.				PR	19-JUL-2001; 2001US-306799P.	
PR	19-JUL-2001; 2001US-306965P.				PR	26-JUL-2001; 2001US-308237P.	
PR	26-JUL-2001; 2001US-308237P.				PR	27-JUL-2001; 2001US-308184P.	
PR	27-JUL-2001; 2001US-310091P.				PR	03-AUG-2001; 2001US-310091P.	

PR 03-AUG-2001; 2001US-310093P.  
 PR 03-AUG-2001; 2001US-310094P.  
 XX  
 PA (INCY- ) INCYTE GENOMICS INC.  
 XX  
 PI Tang YT, Elliott VS, Baughn MR, Yue H, Klammer AA, Hafalia AJA;  
 PI Stewart EA, Honchell CD, Gorvath AE, Forsythe IT, Ding L, Tran UK;  
 PI Griffin JA, Zebardjian Y, Chinn AM, Curtis AL, Borowsky ML;  
 PI Swarnakar A, Burford N, Luo W, Lee EA, Emseling BM, Lee SY;  
 PI Yang J, Khan FA, Ramkumar J, Lu Y, Walla NK, Dugan BM;  
 PI Gietzen KJ, Jones KA;  
 XX  
 DR WPI: 2003-239324/23.  
 DR N-PSDB; ABA00963.  
 XX  
 PT New human proteins associated with cell growth, differentiation and  
 PT death, useful for diagnosing, treating and preventing e.g., cell  
 PT proliferative, developmental, neurological, metabolic, or  
 PT autoimmune/inflammatory disorders -  
 XX  
 PS Claim 1; Page 161-64; 197pp; English.  
 PS  
 XX The sequences given in AAG79926-43 show human proteins associated  
 CC with cell growth, differentiation and death (CGDD)-1 to -18. The human  
 CC CGDD proteins, and the nucleic acids encoding them, are useful for  
 CC diagnosing, treating and preventing cell proliferative disorders (e.g.  
 CC cancer, actinic keratosis, arteriosclerosis, atherosclerosis or  
 CC bursitis), developmental disorders (e.g. renal tubular acidosis,  
 CC anemia, Cushing's syndrome or Duchenne and Becker muscular dystrophy),  
 CC neurological disorders (e.g. epilepsy, stroke, cerebral neoplasms,  
 CC Alzheimer's disease or Pick's disease), autoimmune/inflammatory  
 CC disorders (e.g. AIDS, allergies, asthma, or autoimmune thyroiditis),  
 CC metabolic disorders (e.g. obesity, type II diabetes), reproductive  
 CC disorders (e.g. infertility or disorder of prolactin production), and  
 CC disorders of the placenta (e.g. preeclampsia, chorionicarcoma, abruptio  
 CC placenta or chorioamnionitis). The proteins and nucleic acids are also  
 CC useful in screening for compounds that modulate the activity of CGDD,  
 CC and in creating transgenic animals to model human disease. The  
 CC CGDD-4 protein shows homology to bovine SMCI protein.  
 CC  
 XX  
 SQ Sequence 1254 AA;  
 Query Match 45.6%; Score 41.5; DB 24; Length 1254;  
 Best Local Similarity 56.2%; Pred. No. 4.2e+02;  
 Matches 9; Conservative 4; Mismatches 2; Indels 1; Gaps 1;  
 QY 3 PEOQETV-LDGNLIIR 17  
 Db 633 PERQKTVALDGTLPFK 648  
 RESULT 36  
 ABP70321  
 ID ABP70321 standard; peptide, 28 AA.  
 XX  
 AC ABP70321;  
 XX  
 DT 07-APR-2003 (first entry)  
 XX  
 DB Tryptic peptide of heat shock protein 60 of Chlamydia trachomatis.  
 XX  
 KW Microorganism; mass spectroscopy; biomarker; heat shock protein 60;  
 KW Hsp60.  
 XX  
 OS Chlamydia trachomatis.  
 XX  
 PN WO200295416-A2.  
 XX  
 PD 28-NOV-2002.  
 PD  
 PF 22-MAY-2002; 2002WO-GB02208.  
 PF  
 PR 22-MAY-2001; 2001GB-0012428.  
 PR

XX  
 PA (MINA ) UK SEC FOR DEFENCE.  
 XX  
 PI Tibball RW, Deepeyroux D;  
 XX  
 DR WPI: 2003-129474/12.  
 XX  
 PT New biomarker comprising species homologues derived from the majority  
 PT of species in at least two groups of microorganisms that are  
 PT structurally similar, useful for identifying microorganisms -  
 XX  
 PS Example 2; Page 16; 35pp; English.  
 PS  
 XX The specification describes a method of rapidly identifying unknown  
 CC microorganisms by means of mass spectroscopy of biomarkers. The  
 CC biomarkers comprise species homologues derived from the majority of  
 CC species in at least two groups of microorganisms that are structurally  
 CC similar. The structural similarity allows isolation of the biomarkers  
 CC from different species of microorganism. Each biomarker has an unique  
 CC molecular mass. The biomarker is useful for identifying microorganisms.  
 CC Peptides ABP70319-56 represent a tryptic peptide map of the heat shock  
 CC protein 60 (Hsp60) of Chlamydia trachomatis. Hsp60 is used as a biomarker  
 CC to demonstrate the method of the invention.  
 CC  
 XX  
 SQ Sequence 28 AA;  
 Query Match 45.1%; Score 41; DB 24; Length 28;  
 Best Local Similarity 47.1%; Pred. No. 4.6;  
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
 QY 2 SPEQETVLDGNLIIR 18  
 Db 10 NPERQECVLEDAVLIV 26  
 RESULT 37  
 AAE10209  
 ID AAE10209 standard; Protein; 140 AA.  
 XX  
 AC AAE10209;  
 XX  
 DT 29-NOV-2001 (first entry)  
 XX  
 DE Human bone marrow derived contig protein, SEQ ID NO: 74.  
 XX  
 KW Human; bone marrow; cytokine; cell proliferation; cell differentiation;  
 KW stem cell growth; haematopoiesis regulation; immunoregulatory; anemia;  
 KW Parkinson's disease; Alzheimer's disease; amyotrophic lateral sclerosis;  
 KW infection; HIV; malaria; fungal; cancer; autoimmune disorder; arthritis;  
 KW Crohn's disease; inflammatory bowel disease; leukaemia; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 41  
 FT  
 XX  
 PN WO200166558-A1.  
 XX  
 PD 13-SEP-2001.  
 PD  
 PF 07-MAR-2001; 2001WO-US07274.  
 PF  
 PR 07-MAR-2000; 2000US-0519705.  
 PR 19-MAY-2000; 2000US-0574454.  
 PR 14-JUL-2000; 2000US-0616847.  
 PR 30-NOV-2000; 2000US-0250583.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 XX Ford JE, Boyle BJ, Tang YT, Liu C, Zhou P, Asundi V, Zhang J;  
 PI Drmanac RT;  
 PI

DR WP1; 2001-565565/63.  
 DR N-PSDB; AAD17390.  
 XX  
 PT Isolated polynucleotide encoding bone marrow derived polypeptides  
 PT useful for treating, e.g., Parkinson's, Alzheimer's, cancer, arthritis,  
 PT Crohn's disease, and inflammatory bowel disease -  
 XX  
 PS Claim 10; Page 113; 1699p; English.  
 XX  
 CC The invention relates to novel human bone marrow polynucleotides and  
 CC polypeptides. The polypeptides of the invention may exhibit various  
 CC activities e.g. cytokine and cell proliferation/differentiation/  
 CC stem cell growth, haematopoiesis regulation and immune stimulation/  
 CC suppression. The polynucleotides and polypeptides of the invention may  
 CC be used to detect bone marrow cells, for treating, e.g., Parkinson's  
 CC disease, Alzheimer's disease, anaemia, amyotrophic lateral sclerosis,  
 CC infections such as HIV, malaria and fungal infections, cancer, autoimmune  
 CC disorders, arthritis, Crohn's disease, inflammatory bowel disease, and  
 CC leukaemia. The polynucleotide may also be used as markers for tissues in  
 CC which the corresponding protein is preferentially expressed; as molecular  
 CC weight markers and as probes to hybridise and discover novel, related DNA  
 CC sequences. The polypeptide may be used in assays to determine biological  
 CC activity, to raise antibodies or illicit an immune response and as a  
 CC reagent in assays. The present sequence is a human bone marrow derived  
 CC contig protein of the invention.  
 CC  
 XX  
 SQ Sequence 140 AA;  
 Query Match 45.1%; Score 41; DB 22; Length 140;  
 Best Local Similarity 53.8%; Pred. No. 34;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 3 PEOGETVLDDGNTLI 15  
 Db 28 PPEOETRIDGNTI 40  
 RESULT 38  
 ID AAU18354 standard; Protein; 189 AA.  
 XX  
 AC AAU18354;  
 XX  
 DT 21-NOV-2001 (first entry)  
 XX  
 DE Human endocrine polypeptide SEQ ID No 309.  
 XX  
 KW Endocrine protein; human; mouse; rabbit; goat; horse; food additive;  
 KW cat; dog; chicken; sheep; immunosuppressive; antiarthritic; vasotrophic;  
 KW antirheumatic; antiproliferative; cytostatic; cardiast; neuroprotective;  
 KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;  
 KW ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasm;  
 KW hyperproliferative disorder; breast; liver; cardiovascular disorder;  
 KW cerebrovascular disorder; nervous system disorder; bacterial infection;  
 KW fungal infection; viral infection; ocular disorder; endocrine disorder;  
 KW gastrointestinal disorder; renal disorder; respiratory disorder;  
 KW wound healing; skin aging; organ transplantation; food preservative;  
 KW tissue regeneration; anti-infertility.  
 XX  
 OS Homo sapiens.  
 PN WO200155364-A2.  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01308.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 26-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225477.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226688.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 06-SEP-2000; 2000US-0230439.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234937.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.

CC are useful in the diagnosis, treatment and prevention of various types of  
CC disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,  
CC chickens or sheep. A pathological condition can be determined by  
CC determining the presence or absence of a mutation in an endocrine  
CC polynucleotide. The treatable disorders include autoimmune diseases such  
CC as rheumatoid arthritis, hyperproliferative disorders such as neoplasms  
CC of the breast or liver, cardiovascular disorders such as cardiac arrest,  
CC cerebrovascular disorders such as cerebral ischaemia, nervous system  
CC disorders such as Alzheimer's disease, infections caused by bacteria,  
CC viruses and fungi, ocular disorders such as corneal infection, endocrine  
CC disorders such as premature labour and infertility, gastrointestinal  
CC disorders such as Crohn's disease, renal disorders such as  
CC glomerulonephritis and respiratory disorders such as asthma. The  
CC polypeptides can also be used to aid wound healing, to prevent skin aging  
CC due to sunburn, to maintain organs before transplantation, to regenerate  
CC tissues and in chemotaxis. The polypeptides can also be used as a food  
CC additive or preservative to increase or decrease storage capabilities.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

Query Match 45.1%; Score 41; DB 22; Length 189;  
Best Local Similarity 46.7%; Pred. No. 49;  
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 3 PROEYVLDGNLIR 17  
| : | : | : | : | : | :  
Db 161 PIEQRTVVDEGPVR 175

RESULT 39  
AAV37099  
ID AAV37099 standard; Protein; 258 AA.  
AC AAV37099;  
XX  
DT 07-OCT-1999 (first entry)  
XX  
XX  
DE Amino acid sequence of a Chlamydia trachomatis protein.  
XX  
XX  
KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
KW paratrachoma; inclusion conjunctivitis; genital disease; perithatitis;  
KW nongonococcal urethritis; epididymitis; cervicitis; balanitis;  
KW Bartholinitis; pneumopathy; venereal lymphogranulomatosis.  
XX  
OS Chlamydia trachomatis.  
XX  
PM WO9928475-A2.  
XX  
PD 10-JUN-1999.  
XX  
PF 27-NOV-1998; 98WO-IB01939.  
XX  
PR 04-NOV-1998; 98US-0107077.  
PR 28-NOV-1997; 97FR-0015041.  
PR 17-DEC-1997; 97FR-0016034.  
XX  
PA (GIST ) GENSET.  
XX  
PI Griffais R;  
XX  
DR WPI, 1999-371125/31.  
XX  
PT  
XX Genome sequence of Chlamydia trachomatis  
PS Disclosure; Page 899-900; 175pp; English.

AAV36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
of Chlamydia trachomatis (see A201425). The polypeptides can be used as  
vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
can also be used to control growth of the microorganism. Chlamydia  
trachomatis is responsible for a large number of diseases, e.g. eye  
diseases such as conventional trachoma, nonendemic trachoma,



The present invention relates to the genome sequence of *Listeria monocytogenes* EGD-e (see AB030341). The genome sequence and fragments of it are useful for selecting probes and primers for detecting genes in *L. monocytogenes* and related organisms, and for studying genetic polymorphisms and other genomes. The present sequence is a protein encoded by the genome sequence of the present invention. Proteins expressed from the genome sequence are useful for raising specific antibodies, identification of *L. monocytogenes* and related organisms, and for biosynthesis and biodegradation, especially biosynthesis of Vitamin B12. The genome sequence and proteins encoded by it are also useful for selecting compounds that regulate gene expression and cell replication and modulate *L. monocytogenes*-related diseases. In addition, the genome sequence and proteins encoded by it are useful in pharmaceutical and vaccines compositions for the treatment or prevention of infections by *L.*

Search completed: October 10, 2003, 17:22:30  
Job time : 84 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: October 10, 2003, 17:21:04 ; Search time 29 Seconds  
(without alignments)  
26.262 Million cell updates/sec

Title: US-09-991-795-1

Sequence: 1 KSPQQQETVDGNLIRY 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A COMB .pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B COMB .pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A COMB .pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B COMB .pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PTUS COMB .pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	47.3	946	US-09-074-579-3	Sequence 3, Appli
2	43	47.3	946	US-09-388-774-3	Sequence 3, Appli
3	43	47.3	946	US-09-546-153-1	Sequence 1, Appli
4	42	46.2	544	US-09-198-452A-153	Sequence 153, App
5	42	46.2	652	US-09-347-833-2	Sequence 2, Appli
6	42	46.2	858	US-08-265-628-2	Sequence 2, Appli
7	41.5	45.6	353	US-09-134-001C-4640	Sequence 4640, Ap
8	41	45.1	224	US-09-328-352-5302	Sequence 5302, Ap
9	40.5	44.5	224	US-08-871-5728-9	Sequence 9, Appli
10	40.5	44.5	434	US-08-328-256-11	Sequence 11, Appli
11	40.5	44.5	436	US-08-307-588-2	Sequence 2, Appli
12	40.5	44.5	436	US-07-971-834-2	Sequence 12, Appli
13	40.5	44.5	496	US-08-328-256-12	Sequence 12, Appli
14	40.5	44.5	557	US-08-328-256-10	Sequence 10, Appli
15	40.5	44.5	557	US-08-471-454-2	Sequence 2, Appli
16	40.5	44.5	557	US-08-466-974-2	Sequence 2, Appli
17	40.5	44.5	557	US-08-471-453-2	Sequence 2, Appli
18	40.5	44.5	557	US-08-307-588-4	Sequence 4, Appli
19	40.5	44.5	557	US-07-971-834-4	Sequence 4, Appli
20	40	44.0	520	US-09-752-165-1	Sequence 1, Appli
21	40	44.0	544	US-08-467-822-33	Sequence 33, Appli
22	40	44.0	544	US-08-432-697-33	Sequence 33, Appli
23	40	44.0	544	US-08-466-248-13	Sequence 33, Appli
24	40	44.0	654	US-08-487-890A-12	Sequence 12, Appli
25	40	44.0	654	US-08-478-435-12	Sequence 12, Appli
26	40	44.0	654	US-08-337-483-12	Sequence 12, Appli
27	40	44.0	654	US-08-478-373-12	Sequence 12, Appli

28	40	44.0	654	3	US-08-474-671-12	Sequence 12, Appli
29	40	44.0	654	3	US-08-483-577A-12	Sequence 12, Appli
30	40	44.0	654	3	US-08-897-438-12	Sequence 12, Appli
31	40	44.0	654	4	US-08-637-654-12	Sequence 12, Appli
32	40	44.0	654	4	US-08-649-518-12	Sequence 12, Appli
33	40	44.0	769	3	US-08-434-000A-10	Sequence 10, Appli
34	40	44.0	769	4	US-09-312-157-10	Sequence 10, Appli
35	40	44.0	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
36	39	42.9	138	4	US-09-107-532A-4089	Sequence 4089, Ap
37	39	42.9	430	1	US-08-785-076-3	Sequence 3, Appli
38	39	42.9	430	4	US-09-018-824-3	Sequence 3, Appli
39	39	42.9	432	1	US-08-785-076-2	Sequence 2, Appli
40	39	42.9	432	4	US-09-018-824-2	Sequence 2, Appli
41	38.5	42.3	260	4	US-09-252-991A-21181	Sequence 21181, A
42	38.5	42.3	323	1	US-07-913-107-2	Sequence 2, Appli
43	38.5	42.3	323	1	US-08-459-201-2	Sequence 2, Appli
44	38.5	42.3	323	1	US-08-281-248-2	Sequence 2, Appli
45	38.5	42.3	323	5	PCT-US93-06421-1	Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-09-074-579-3  
Sequence 3, Application US/09074579  
Patent No. 6001596  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Guegler, Karl J.  
APPLICANT: Patterson, Chandra  
TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE  
TITLE OF INVENTION: INHIBITOR HEAVY CHAIN PRECURSOR  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
OPERATING SYSTEM: DOS  
SOFTWARE: Word Perfect 6.1/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/074,579  
FILING DATE: HERewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Cerrone, Michael C  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0505 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 946 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GENBANK  
CLONE: g133985  
US-09-074-579-3  
Query Match 47.3%, Score 43, DB 3, Length 946;

Best Local Similarity 43.8%; Pred. No. 59;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 3 PEOQETVLDGNLIIRY 18  
Db 262 PSCRETAVDGELVLY 277

## RESULT 2

US-09-388-774-3  
; Sequence 3, Application US/09388774  
; Patent No. 6228991  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Patterson, Chandra  
; TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE  
; TITLE OF INVENTION: INHIBITOR HEAVY CHAIN PRECURSOR  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: Word Perfect 6.1/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/388,774  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/074,579  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cerrione, Michael C  
; REGISTRATION NUMBER: 39,132  
; REFERENCE/DOCKET NUMBER: PF-0505 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650-855-0555  
; TELEFAX: 650-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 946 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GENBANK  
; CLONE: g133985  
; US-09-388-774-3

## Query Match

Best Local Similarity 47.3%; Score 43; DB 3; Length 946;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 3 PEOQETVLDGNLIIRY 18  
Db 262 PSCRETAVDGELVLY 277

## RESULT 3

US-09-546-153-1  
; Sequence 1, Application US/09546153  
; Patent No. 6448054  
; GENERAL INFORMATION:  
; APPLICANT: Poznansky, Mark C.  
; APPLICANT: Luster, Andrew D.

; APPLICANT: Scadden, David T.  
; TITLE OF INVENTION: PURPOSEFUL MOVEMENT OF HUMAN MIGRATORY  
; TITLE OF INVENTION: CELLS AWAY FROM AN AGENT SOURCE  
; FILE REFERENCE: M0765/7025/ERG/KA  
; CURRENT APPLICATION NUMBER: US/09/546,153  
; CURRENT FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: U.S. 60/128,272  
; PRIOR FILING DATE: 1999-04-08  
; PRIOR APPLICATION NUMBER: U.S. 60/168,952  
; PRIOR FILING DATE: 1999-12-03  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 946  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
US-09-546-153-1

Qy 3 PEOQETVLDGNLIIRY 18  
Db 262 PSCRETAVDGELVLY 277

## Query Match

Best Local Similarity 47.3%; Score 43; DB 4; Length 946;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

US-09-198-452A-153  
; Sequence 153, Application US/09198452A  
; Patent No. 6559294  
; GENERAL INFORMATION:  
; APPLICANT: Griffiths, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; CURRENT FILING DATE: 1998-11-24  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 153  
; LENGTH: 544  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: 1..544  
; OTHER INFORMATION: Xaa=unknown or other  
US-09-198-452A-153

Query Match 46.2%; Score 42; DB 4; Length 544;  
Best Local Similarity 52.9%; Pred. No. 45;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SPOQETVLDGNLIIRY 18  
Db 207 NPETOQCVLDELALILY 223

## Query Match

US-09-347-833-2  
; Sequence 2, Application US/09347833  
; Patent No. 6294658  
; GENERAL INFORMATION:  
; APPLICANT: Ramodu, Layo O.  
; APPLICANT: Odell, Joan T.  
; TITLE OF INVENTION: Factors Involved in Gene Expression  
; FILE REFERENCE: BB-1172  
; CURRENT APPLICATION NUMBER: US/09/347,833  
; CURRENT FILING DATE: 1998-07-02  
; EARLIER APPLICATION NUMBER: 60/092,415  
; EARLIER FILING DATE: July 10, 1998  
; NUMBER OF SEQ ID NOS: 11

SOFTWARE: Microsoft Office 97  
SEQ ID NO 2  
LENGTH: 652  
TYPE: PRT  
ORGANISM: Zea mays  
US-09-947-833-2

Query Match 46.2%; Score 42; DB 3; Length 652;  
Best Local Similarity 61.5%; Pred. No. 56;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 SPEQOETVLDGNTL 14  
:|||||:|  
Db 565 NPEQOETLIGENL 577

RESULT 6  
US-08-265-628-2  
Sequence 2, Application US/08265628  
Patent No. 5821094  
GENERAL INFORMATION:  
APPLICANT: Rothstein, Steven J.  
APPLICANT: Goring, Daphne  
TITLE OF INVENTION: S-LOCUS RECEPTOR KINASE GENE IN A  
TITLE OF INVENTION: SELF-INCOMPATIBLE BRASSICA NAPUS LINE  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McAndrews, Held & Malloy, Ltd.  
STREET: 500 W. Madison St. Suite 3400  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60661  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/265,628  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/959,945  
FILING DATE: US 07/847,564  
APPLICATION NUMBER: 03-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pochopien Ph.D., Donald J.  
REGISTRATION NUMBER: 32,167  
TELEPHONE: 312-707-8889  
TELEFAX: 312-707-9155  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 858 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-265-628-2

Query Match 46.2%; Score 42; DB 2; Length 858;  
Best Local Similarity 44.4%; Pred. No. 78;  
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 KSPEQOETVLDGNTLIRY 18  
:|||:|:|:|  
Db 130 RSPVALLLENGNVIRY 147

RESULT 7  
US-09-134-001C-4640  
Sequence 4640, Application US/09134001C

Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4640  
LENGTH: 353  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4640

Query Match 45.6%; Score 41.5; DB 4; Length 353;  
Best Local Similarity 55.0%; Pred. No. 32;  
Matches 11; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

QY 2 SPEQOETVLD-----GNTLII 16  
:|||||:|  
Db 261 SPMDQETVLDVKKTGRLLIY 280

RESULT 8  
US-09-328-352-5302  
Sequence 5302, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 5302  
LENGTH: 224  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-5302

Query Match 45.1%; Score 41; DB 4; Length 224;  
Best Local Similarity 47.1%; Pred. No. 22;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 SPEQOETVLDGNTLIRY 18  
:|||:|:|:|  
Db 66 SPDIDESVHDELVHQY 82

RESULT 9  
US-08-871-572B-9  
Sequence 9, Application US/08871572B  
Patent No. 6287853  
GENERAL INFORMATION:  
APPLICANT: Peetka, Sidney  
APPLICANT: Kosenko, Serguei  
APPLICANT: Sol, Jaemog  
APPLICANT: Donnelly, Robert  
APPLICANT: Mariano, Thomas  
APPLICANT: Cook, Jeffrey  
APPLICANT: Emmanuel, Stuart  
APPLICANT: Schwartz, Barbara  
TITLE OF INVENTION: Accessory Factor for Interferon Gamma  
TITLE OF INVENTION: and Its Receptor  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Richard R. Muccino

STREET: 758 Springfield Avenue  
CITY: Summit  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07901  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/871,572B  
FILING DATE: 9-JUNE-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Muccino, Richard R.  
REGISTRATION NUMBER: 32,538  
REFERENCE/DOCKET NUMBER: UMD1-011  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908) 273-4988  
TELEFAX: (908) 273-4679  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 224 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHEITICAL: NO  
ANTI-SENSE: NO  
US-08-871-572B-9

Query Match 44.5%; Score 40.5; DB 3; Length 224;  
Best Local Similarity 42.1%; Pred. No. 27;  
Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

QY 1 KSPQOQT-VLDGTLIRY 18  
Db 31 KSPQKVEVDIIDNFILRW 49

RESULT 10  
US-08-328-256-11  
Sequence 11, Application US/08328256  
Patent No. 5643749  
GENERAL INFORMATION:  
APPLICANT: REVEL, Michel  
APPLICANT: ABRAMOVICH, Carolina  
APPLICANT: RATOVIYTSKI, Edward  
TITLE OF INVENTION: SOLUBLE INTERFERON ALPHA-RECEPTOR, ITS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROWDY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/328,256  
FILING DATE: 24-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 107378  
FILING DATE: 24-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: REVEL=13

REFERENCE/DOCKET NUMBER: 25,618  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 434 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-328-256-11

Query Match 44.5%; Score 40.5; DB 1; Length 434;  
Best Local Similarity 42.1%; Pred. No. 61;  
Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

QY 1 KSPQOQT-VLDGTLIRY 18  
Db 31 KSPQKVEVDIIDNFILRW 49

RESULT 11  
US-08-307-588-2  
Sequence 2, Application US/08307588  
Patent No. 5919453  
GENERAL INFORMATION:  
APPLICANT: BENOIT, Patrick  
APPLICANT: MEYER, Francois  
APPLICANT: MAGUIRE, Deborah  
APPLICANT: PLAVEC, Ivan  
TITLE OF INVENTION: MONOCLONAL ANTIBODIES AGAINST THE INTERFERON  
RECEPTOR, WITH NEUTRALIZING ACTIVITY AGAINST TYPE I  
TITLE OF INVENTION: INTERFERON  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
ZIP: 20007  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (ERO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/307,588  
FILING DATE: 05-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP93/00770  
FILING DATE: 30-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 92400902.0  
FILING DATE: 31-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: SAYE, Bernhard D.  
REGISTRATION NUMBER: 28,665  
REFERENCE/DOCKET NUMBER: 17283/117/GUPL  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 436 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-307-588-2

Query Match 44.5%; Score 40.5; DB 2; Length 436;

Best Local Similarity 42.1%; Pred. No. 61;  
Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;  
QY 1 KSPQOQT-VLDGULIRY 18  
|||:|::|:|:|:  
Db 31 KSPQKVEVDIIDDNFILRW 49

RESULT 12  
US-07-971-834-2  
; Sequence 2, Application US/07971834  
; Patent No. 6475983  
; GENERAL INFORMATION:  
; APPLICANT: EID, Pierre  
; APPLICANT: GRESSER, Ion  
; APPLICANT: LUTERALL, Georges  
; APPLICANT: MEYER, Francois  
; APPLICANT: MOGENSEN, Knud E.  
; APPLICANT: TOVEY, Michael  
; APPLICANT: UZE, Gilles  
; TITLE OF INVENTION: WATER-SOLUBLE POLYPEPTIDES HAVING HIGH  
; TITLE OF INVENTION: AFFINITY FOR INTERFERONS ALPHA AND BETA  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/971.834  
; FILING DATE: 17-FEB-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/FR91/00318  
; FILING DATE: 17-APR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: EID=1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 436 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-971-834-2  
Query Match 44.5%; Score 40.5; DB 4; Length 436;  
Best Local Similarity 42.1%; Pred. No. 61;  
Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;  
QY 1 KSPQOQT-VLDGULIRY 18  
|||:|::|:|:|:  
Db 31 KSPQKVEVDIIDDNFILRW 49

APPLICANT: ABRAMOVICH, Carolina  
; APPLICANT: RATOVIITSKI, Edward  
; TITLE OF INVENTION: SOLUBLE INTERFERON ALPHA-RECEPTOR, ITS  
; TITLE OF INVENTION: PREPARATION AND USE  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/328,256  
; FILING DATE: 24-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: IL 107378  
; FILING DATE: 24-OCT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.  
; REGISTRATION NUMBER: REVEL=13  
; REFERENCE/DOCKET NUMBER: 25,618  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; TELEX: 248633  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 496 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-328-256-12  
Query Match 44.5%; Score 40.5; DB 1; Length 496;  
Best Local Similarity 42.1%; Pred. No. 72;  
Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;  
QY 1 KSPQOQT-VLDGULIRY 18  
|||:|::|:|:|:  
Db 31 KSPQKVEVDIIDDNFILRW 49

RESULT 14  
US-08-328-256-10  
; Sequence 10, Application US/08328256  
; Patent No. 5643749  
; GENERAL INFORMATION:  
; APPLICANT: REVEL, Michel  
; APPLICANT: ABRAMOVICH, Carolina  
; APPLICANT: RATOVIITSKI, Edward  
; TITLE OF INVENTION: SOLUBLE INTERFERON ALPHA-RECEPTOR, ITS  
; TITLE OF INVENTION: PREPARATION AND USE  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/328,256  
FILING DATE: 24-OCT-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 107378  
FILING DATE: 24-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, ROGER L.  
REGISTRATION NUMBER: REVEL-13  
REFERENCE/DOCKET NUMBER: 25,618  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 557 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-328-256-10

Query Match 44.5% Score 40.5; DB 1; Length 557;  
Best Local Similarity 42.1%; Pred. No. 83;  
Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

Qy 1 KSPQOET-VLDGNLIIRY 18  
Db 31 KSPQKVEVDIIDNFIIRW 49

RESULT 15  
US-08-471-454-2  
Sequence 2, Application US/08471454

PATENT No. 5731169  
GENERAL INFORMATION:  
APPLICANT: MOGENSEN, Knud E.  
APPLICANT: UZE, Gilles  
APPLICANT: LUTFALLA, Georges  
APPLICANT: GRESSER, Ion  
TITLE OF INVENTION: CDNA FRAGMENT CODING FOR THE GENE FOR  
TITLE OF INVENTION: THE ALPHA INTERFERON RECEPTOR AND PROCESS FOR THE  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHAYE P.C.  
STREET: 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,454  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/900,642  
FILING DATE: 15-JUN-1992  
APPLICATION NUMBER: FR 89/13770  
FILING DATE: 20-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: BYRNE, THOMAS E.  
REGISTRATION NUMBER: 32,205  
REFERENCE/DOCKET NUMBER: 960-7  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
TELEX: 200797 NIXN UR

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 557 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-471-454-2

Query Match 44.5% Score 40.5; DB 1; Length 557;  
Best Local Similarity 42.1%; Pred. No. 83;  
Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

Qy 1 KSPQOET-VLDGNLIIRY 18  
Db 31 KSPQKVEVDIIDNFIIRW 49

RESULT 16  
US-08-466-974-2  
Sequence 2, Application US/08466974

PATENT No. 5861258  
GENERAL INFORMATION:  
APPLICANT: MOGENSEN, Knud E.  
APPLICANT: UZE, Gilles  
APPLICANT: LUTFALLA, Georges  
APPLICANT: GRESSER, Ion  
TITLE OF INVENTION: CDNA FRAGMENT CODING FOR THE GENE FOR  
TITLE OF INVENTION: THE ALPHA INTERFERON RECEPTOR AND PROCESS FOR THE  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHAYE P.C.  
STREET: 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,974  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/900,642  
FILING DATE: 15-JUN-1992  
APPLICATION NUMBER: FR 89/13770  
FILING DATE: 20-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: BYRNE, THOMAS E.  
REGISTRATION NUMBER: 32,205  
REFERENCE/DOCKET NUMBER: 960-7  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
TELEX: 200797 NIXN UR  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 557 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-466-974-2

Query Match 44.5% Score 40.5; DB 2; Length 557;  
Best Local Similarity 42.1%; Pred. No. 83;  
Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;



QY 1 KSPKQET-VLDGNLIRY 18  
|||:|:|:|:|:  
Db 31 KSPKQEVVDIIDDNFILRW 49

## RESULT 17

US-08-471-453-2  
; Sequence 2, Application US/08471453  
; Patent No. 5886153  
; GENERAL INFORMATION:  
; APPLICANT: MOGENSEN, Knud E.  
; APPLICANT: UZE, Gilles  
; APPLICANT: LUTFALLA, Georges  
; APPLICANT: GRESSER, Ion  
; TITLE OF INVENTION: CDNA FRAGMENT CODING FOR THE GENE FOR  
; TITLE OF INVENTION: THE ALPHA INTERFERON RECEPTOR AND PROCESS FOR THE  
; TITLE OF INVENTION: PREPARATION OF THE CORRESPONDING PROTEIN  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHAYE P.C.  
; STREET: 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: U.S.A.  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/471.453  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/900,642  
; FILING DATE: 15-JUN-1992  
; APPLICATION NUMBER: FR 89/13770  
; FILING DATE: 20-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BYRNE, THOMAS E.  
; REGISTRATION NUMBER: 32,205  
; REFERENCE/DOCKET NUMBER: 960-7  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 816-4000  
; TELEFAX: (703) 816-4100  
; TELEX: 200797 NIXN UR  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 557 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-471-453-2

Query Match 44.5%; Score 40.5; DB 2; Length 557;

Best Local Similarity 42.1%; Pred. No. 83;  
Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

QY 1 KSPKQET-VLDGNLIRY 18  
|||:|:|:|:|:  
Db 31 KSPKQEVVDIIDDNFILRW 49

## RESULT 18

US-08-307-588-4  
; Sequence 4, Application US/08307588  
; Patent No. 5919453  
; GENERAL INFORMATION:  
; APPLICANT: BENOIT, Patrick  
; APPLICANT: MEYER, Francois  
; APPLICANT: MAGUIRE, Deborah

APPLICANT: PLAVEC, Ivan  
; APPLICANT: TOVEY, Michael G.  
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES AGAINST THE INTERFERON  
; TITLE OF INVENTION: RECEPTOR, WITH NEUTRALIZING ACTIVITY AGAINST TYPE I  
; TITLE OF INVENTION: INTERFERON  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; ZIP: 20007

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent in Release #1.0, Version #1.25 (BEO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/307,588  
; FILING DATE: 05-DEC-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP93/00770  
; FILING DATE: 30-MAR-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 92400902.0  
; FILING DATE: 31-MAR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saxe, Bernhard D.  
; REGISTRATION NUMBER: 28,665  
; REFERENCE/DOCKET NUMBER: 17283/117/CUPL  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 557 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-307-588-4

Query Match 44.5%; Score 40.5; DB 2; Length 557;  
Best Local Similarity 42.1%; Pred. No. 83;  
Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

QY 1 KSPKQET-VLDGNLIRY 18  
|||:|:|:|:|:  
Db 31 KSPKQEVVDIIDDNFILRW 49

RESULT 19  
US-07-971-834-4  
; Sequence 4, Application US/07971834  
; Patent No. 6475983  
; GENERAL INFORMATION:  
; APPLICANT: EID, Pierre  
; APPLICANT: GRESSER, Ion  
; APPLICANT: LUTFALLA, Georges  
; APPLICANT: MEYER, Francois  
; APPLICANT: MOGENSEN, Knud E.  
; APPLICANT: TOVEY, Michael

APPLICANT: UZE, Gilles  
; TITLE OF INVENTION: WATER-SOLUBLE POLYPEPTIDES HAVING HIGH  
; TITLE OF INVENTION: AFFINITY FOR INTERFERONS ALPHA AND BETA  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/971,834  
FILING DATE: 17-FEB-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR91/00318  
FILING DATE: 17-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BROWDY, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: EID=1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 557 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-971-834-4

Query Match 44.5% Score 40.5; DB 4; Length 557;  
Best Local Similarity 42.1% Pred. No. 83;  
Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

QY 1 KSPQOETVLDGNLIIRY 18  
|||::|||::|||:  
Db 31 KSPQKVEVDIIDDNFIILRW 49

RESULT 20  
US-09-752-165-1  
Sequence 1, Application US/09752165  
Patent No. 6451583  
GENERAL INFORMATION:  
APPLICANT: Shuman, Stewart  
APPLICANT: Ho, C. Kiong  
TITLE OF INVENTION: Pharmacological Targeting Of mRNA Cap Formation  
TITLE OF INVENTION: For Treatment Of Parasitic Infections  
FILE REFERENCE: D6388  
CURRENT APPLICATION NUMBER: US/09/752,165  
CURRENT FILING DATE: 2000-12-29  
NUMBER OF SEQ ID NOS: 104  
SEQ ID NO 1  
LENGTH: 520  
TYPE: PRT  
ORGANISM: Plasmodium falciparum  
FEATURE:  
NAME/KEY: PEPTIDE  
OTHER INFORMATION: Plasmodium falciparum RNA guanylyltransferase Pgt1  
US-09-752-165-1

Query Match 44.0% Score 40; DB 4; Length 520;  
Best Local Similarity 57.1% Pred. No. 93;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 SPOQETVLDGNLI 15  
|::|||::|||:  
Db 100 SKQOQIUDGRLV 113

RESULT 21  
US-08-467-822-33  
Sequence 33, Application US/08467822  
Patent No. 5843460  
GENERAL INFORMATION:  
APPLICANT: Labigne, Agnes

APPLICANT: Sauerbaum, Sebastien  
APPLICANT: Ferrero, Richard L.  
APPLICANT: Thiberge, Jean-Michel  
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST  
TITLE OF INVENTION: HELICOBLACTER INFECTION, POLYPEPTIDES FOR USE IN THE  
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESS: Dunnet  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,822  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/447,177  
FILING DATE: 19-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/432,697  
FILING DATE: 02-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03495.0137-02000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 544 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-467-822-33

Query Match 44.0% Score 40; DB 2; Length 544;  
Best Local Similarity 47.1% Pred. No. 98;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 SPOQETVLDGNLIIRY 18  
|||::|||::|||:  
Db 207 NPETQECVLEBALVLIY 223

RESULT 22  
US-08-432-697-33  
Sequence 33, Application US/08432697  
Patent No. 6248330  
GENERAL INFORMATION:  
APPLICANT: Labigne, Agnes  
APPLICANT: Sauerbaum, Sebastien  
APPLICANT: Ferrero, Richard L.  
APPLICANT: Thiberge, Jean-Michel  
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST  
TITLE OF INVENTION: HELICOBLACTER INFECTION, POLYPEPTIDES FOR USE IN THE  
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/432,697  
FILING DATE: 02-MAY-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03495.0137-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 544 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-432-697-33

Query Match 44.0%; Score 40; DB 3; Length 544;  
Best Local Similarity 47.1%; Pred. No. 98;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SPCQETVLQNLIRY 18  
Db 207 NPTEQCEVLEALVLIY 223

RESULT 23  
US-08-466-248-33  
Sequence 33, Application US/08466248  
Patent No. 6258359  
GENERAL INFORMATION:  
APPLICANT: Labigne, Agnes  
APPLICANT: Sauerbaum, Sebastien  
APPLICANT: Ferrero, Richard L.  
APPLICANT: Thiberge, Jean-Michel  
TITLE OF INVENTION: IMMUNOGENIC COMPOSITIONS AGAINST  
TITLE OF INVENTION: HELICOBACTER INFECTION, POLYPEPTIDES FOR USE IN THE  
TITLE OF INVENTION: COMPOSITIONS, AND NUCLEIC ACID SEQUENCES ENCODING SAID  
TITLE OF INVENTION: POLYPEPTIDES  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/466,248  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/447,177  
FILING DATE: 19-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/432,697  
FILING DATE: 02-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03495.0137-02000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 408-4000  
TELEFAX: (202) 408-4400  
INFORMATION FOR SEQ ID NO: 33:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 544 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-466-248-33

Query Match 44.0%; Score 40; DB 3; Length 544;  
Best Local Similarity 47.1%; Pred. No. 98;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SPCQETVLQNLIRY 18  
Db 207 NPTEQCEVLEALVLIY 223

RESULT 24  
US-08-467-890A-12  
Sequence 12, Application US/08487890A  
Patent No. 5708149  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena  
APPLICANT: Harkness, Robin  
APPLICANT: Schryvers, Anthony  
APPLICANT: Chong, Pele  
APPLICANT: Gray-Owen, Scott  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Murdin, Andrew  
APPLICANT: Klein, Michel  
TITLE OF INVENTION: Transferrin Receptor Genes  
NUMBER OF SEQUENCES: 147  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,890A  
FILING DATE: 07-JUN-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,116  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,968  
FILING DATE: 08-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-466 MTS:jb

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 654 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-487-890A-12

Query Match 44.0%; Score 40; DB 1; Length 654;  
Best Local Similarity 47.4%; Pred. No. 1.2e+02;  
Matches 9; Conservative 4; Mismatches 2; Indels 4; Gaps 1;

OY 1 KSPSEQ---ETVLDGNLI 15  
::|||::|||::|||  
Db 330 QNPENQKSTETTLIDGKLI 348

RESULT 25  
US-08-478-435-12  
Sequence 12, Application US/08478435  
Patent No. 5922323  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena  
APPLICANT: Harkness, Robin  
APPLICANT: Schryvers, Anthony  
APPLICANT: Chong, Pele  
APPLICANT: Gray-Owen, Scott  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Murdin, Andrew  
APPLICANT: Klein, Michel  
TITLE OF INVENTION: Transferrin Receptor Genes  
NUMBER OF SEQUENCES: 147  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/478,435  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/337,483  
FILING DATE: 08-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,116  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,968  
FILING DATE: 08-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-462 MIS:Vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 654 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
US-08-478-435-12

Query Match 44.0%; Score 40; DB 2; Length 654;  
Best Local Similarity 47.4%; Pred. No. 1.2e+02;  
Matches 9; Conservative 4; Mismatches 2; Indels 4; Gaps 1;

OY 1 KSPSEQ---ETVLDGNLI 15  
::|||::|||::|||  
Db 330 QNPENQKSTETTLIDGKLI 348

RESULT 26  
US-08-337-483-12  
Sequence 12, Application US/08337483  
Patent No. 5922562  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena  
APPLICANT: Harkness, Robin  
APPLICANT: Schryvers, Anthony  
APPLICANT: Chong, Pele  
APPLICANT: Gray-Owen, Scott  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Murdin, Andrew  
APPLICANT: Klein, Michel  
TITLE OF INVENTION: Transferrin Receptor Genes  
NUMBER OF SEQUENCES: 147  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/337,483  
FILING DATE: 08-NOV-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-410 MIS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 654 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-337-483-12

Query Match 44.0%; Score 40; DB 2; Length 654;  
Best Local Similarity 47.4%; Pred. No. 1.2e+02;  
Matches 9; Conservative 4; Mismatches 2; Indels 4; Gaps 1;

OY 1 KSPSEQ---ETVLDGNLI 15  
::|||::|||::|||  
Db 330 QNPENQKSTETTLIDGKLI 348

RESULT 27  
US-08-478-373-12  
Sequence 12, Application US/08478373  
Patent No. 5922841  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena

```

; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,373
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-463 MTS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 654 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-478-373-12

Query Match          44.0%; Score 40; DB 2; Length 654;
Best Local Similarity 47.4%; Pred. No. 1.2e+02;
Matches 9; Conservative 4; Mismatches 2; Indels 4; Gaps 1;

QY      1 KSPEQQ---ETVLDNLI 15
       :|||  |||  |||  |||
Db      330 QNPENQKLTETTLIDKLI 348

RESULT 28
US-08-474-671-12
; Sequence 12, Application US/08474671
; Patent No. 6008326
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
```

```

; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,671
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-465 MTS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 654 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-474-671-12

Query Match          44.0%; Score 40; DB 3; Length 654;
Best Local Similarity 47.4%; Pred. No. 1.2e+02;
Matches 9; Conservative 4; Mismatches 2; Indels 4; Gaps 1;

QY      1 KSPEQQ---ETVLDNLI 15
       :|||  |||  |||  |||
Db      330 QNPENQKLTETTLIDKLI 348

RESULT 29
US-08-483-577A-12
; Sequence 12, Application US/08483577A
; Patent No. 6015688
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 160
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,577A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-511
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 654 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-483-577A-12
```

```
Query Match          44.0%; Score 40; DB 3; Length 654;
Best Local Similarity 47.4%; Pred. No. 1.2e+02;
Matches 9; Conservative 4; Mismatches 2; Indels 4; Gaps 1;
```

```
QY      1 KSPBQ---ETVLDGNLI 15
Db      330 QNPENQKLTETLIDGKLI 348
```

```
RESULT 30
US-08-897-438-12
Sequence 12, Application US/08897438
Patent No. 6262016
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena
APPLICANT: Harkness, Robin
APPLICANT: Schryvers, Anthony
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-ping
APPLICANT: Mordin, Andrew
APPLICANT: Klein, Michel
TITLE OF INVENTION: Transferrin Receptor Genes
NUMBER OF SEQUENCES: 160
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: MSG 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/897,438
FILING DATE:
CLASSIFICATION: 435
```

```
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/483,577
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-720
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 654 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-897-438-12
```

```
Query Match          44.0%; Score 40; DB 3; Length 654;
Best Local Similarity 47.4%; Pred. No. 1.2e+02;
Matches 9; Conservative 4; Mismatches 2; Indels 4; Gaps 1;
```

```
QY      1 KSPBQ---ETVLDGNLI 15
Db      330 QNPENQKLTETLIDGKLI 348
```

```
RESULT 31
US-08-637-654-12
Sequence 12, Application US/08637654
Patent No. 6358727
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M
APPLICANT: Harkness, Robin B
APPLICANT: Schryvers, Anthony B
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-ping
APPLICANT: Mordin, Andrew D
APPLICANT: Klein, Michel H
TITLE OF INVENTION: HAEMOPHILUS TRANSFERRIN RECEPTOR GENES
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: MSG 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/637,654
FILING DATE: 05-AUG-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/CA94/00616
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
```

```

; REFERENCE/DOCKET NUMBER: 1038-595
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 654 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-637-654-12

Query Match          44.0%; Score 40; DB 4; Length 654;
Best Local Similarity 47.4%; Pred. No. 1.2e+02;
Matches 9; Conservative 4; Mismatches 2; Indels 4; Gaps 1;

QY 1 KSPSEQ---ETVLDNLI 15
Db 330 QNPENQKLSTETLIDGKLI 348

RESULT 32
US-08-649-518-12
; Sequence 12, Application US/08649518
; Patent No. 6361779
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Hartness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Mordin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 160
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/649,518
; FILING DATE: 17-MAY-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/483,577
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
```

```

; LENGTH: 654 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-649-518-12

Query Match          44.0%; Score 40; DB 4; Length 654;
Best Local Similarity 47.4%; Pred. No. 1.2e+02;
Matches 9; Conservative 4; Mismatches 2; Indels 4; Gaps 1;

QY 1 KSPSEQ---ETVLDNLI 15
Db 330 QNPENQKLSTETLIDGKLI 348

RESULT 33
US-08-434-000A-10
; Sequence 10, Application US/08434000A
; Patent No. 6046037
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,000A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 1
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; TELEX: SEQUENCE LISTING
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-434-000A-10 Rat Polymunoglobulin Receptor

Query Match          44.0%; Score 40; DB 3; Length 769;
Best Local Similarity 63.6%; Pred. No. 1.5e+02;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 SPEGQETVLDG 12
Db 706 TPTQGTETVLEG 716
```

RESULT 34  
US-09-312-157-10  
; Sequence 10, Application US/09312157  
; Patent No. 6303341  
; GENERAL INFORMATION:  
; APPLICANT: ANDREW C. HIATT, JULIAN  
; K.-C. MA, THOMAS LEHNER  
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION  
; PROTEINS IN PLANTS AND THEIR USES  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/312,157  
; FILING DATE: 14-May-1999  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/434,000  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Guise, Jeffrey W.  
; REGISTRATION NUMBER: 34,613  
; REFERENCE/DOCKET NUMBER: 212/127  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 552-8400  
; TELEFAX: (619) 552-0159  
; TELEX: 67-351  
; SEQUENCE LISTING  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 769 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; DESCRIPTION: Rat Polyclonalglobulin Receptor  
; US-09-312-157-10

Query Match 44.0%; Score 40; DB 4; Length 769;  
Best Local Similarity 63.6%; Pred. No. 1.5e+02;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 SPEQOETVLGD 12  
: : : : : :  
Db 706 TPDQETVLG 716

RESULT 35  
US-09-134-001C-5080  
; Sequence 5080, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964

;; PRIOR FILING DATE: 1997-11-08  
;; PRIOR APPLICATION NUMBER: US 60/055,779  
;; PRIOR FILING DATE: 1997-08-14  
;; NUMBER OF SEQ ID NOS: 5674  
;; SEQ ID NO 5080  
;; LENGTH: 3696  
;; TYPE: PRT  
;; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5080

Query Match 44.0%; Score 40; DB 4; Length 3696;  
Best Local Similarity 53.8%; Pred. No. 1e+03;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 KSPEQOETVLGDN 13  
: : : : : :  
Db 1871 KKAEOQVTLINGN 1883

RESULT 36  
US-09-107-532A-4089  
; Sequence 4089, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arinello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 4089:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 138 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (B) LOCATION 1...138  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4089:  
US-09-107-532A-4089

Query Match 42.9%; Score 39; DB 4; Length 138;  
Best Local Similarity 46.2%; Pred. No. 27;  
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;



QY 4 EQQETVLQGNLII 16  
Db 59 EQQETVLQGNLII 71

RESULT 37  
US-08-785-076-3  
; Sequence 3, Application US/08785076  
; Patent No. 5789217  
; GENERAL INFORMATION:  
; APPLICANT: Hodgson, John  
; TITLE OF INVENTION: No. 5789217el tRNA Synthetase  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-0939  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/785.076  
; FILING DATE: 17-JAN-1997  
; CLASSIFICATION: 514  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: 9601096.2  
; FILING DATE: 19-JAN-1996  
; APPLICATION NUMBER: 9615845.6  
; FILING DATE: 27-JUL-1996  
; APPLICATION NUMBER: 9622617.0  
; FILING DATE: 30-OCT-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gimmi, Edward R  
; REGISTRATION NUMBER: 38,891  
; REFERENCE/DOCKET NUMBER: P1354-3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-270-4478  
; TELEFAX: 610-270-5090  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 430 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-785-076-3

Query Match 42.9%; Score 39; DB 1; Length 430;  
Best Local Similarity 57.1%; Pred. No. 1.1e+02;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 SPEQETVLQGNLI 15  
Db 333 NPENETVLQADLI 346

RESULT 38  
US-09-018-824-3  
; Sequence 3, Application US/09018824  
; Patent No. 6410286  
; GENERAL INFORMATION:  
; APPLICANT: Hodgson, John  
; TITLE OF INVENTION: No. 6410286el tRNA Synthetase  
; NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/018.824  
FILING DATE:  
CLASSIFICATION: 435

PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/785.076  
FILING DATE: 17-JAN-1997  
APPLICATION NUMBER: 9601096.2  
FILING DATE: 19-JAN-1996  
APPLICATION NUMBER: 9615845.6  
FILING DATE: 27-JUL-1996  
APPLICATION NUMBER: 9622617.0  
FILING DATE: 30-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimmi, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P1354-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 430 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-018-824-3

Query Match 42.9%; Score 39; DB 4; Length 430;  
Best Local Similarity 57.1%; Pred. No. 1.1e+02;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 SPEQETVLQGNLI 15  
Db 333 NPENETVLQADLI 346

RESULT 39  
US-08-785-076-2  
; Sequence 2, Application US/08785076  
; Patent No. 5789217  
; GENERAL INFORMATION:  
; APPLICANT: Hodgson, John  
; TITLE OF INVENTION: No. 5789217el tRNA Synthetase  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-0939  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/785,076  
FILING DATE: 17-JAN-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9601096.2  
FILING DATE: 19-JAN-1996  
APPLICATION NUMBER: 9615845.6  
FILING DATE: 27-JUL-1996  
APPLICATION NUMBER: 9622617.0  
FILING DATE: 30-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimmil, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P31354-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 432 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-785-076-2

Query Match 42.9%; Score 39; DB 1; Length 432;  
Best Local Similarity 57.1%; Pred. No. 1.1e+02;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SPOQETVLDGNLI 15  
Db 335 NPENETVLCADLI 348

RESULT 40  
US-09-018-824-2  
Sequence 2, Application US/09018824  
Patent No. 6410286  
GENERAL INFORMATION:  
APPLICANT: Hodgson, John  
APPLICANT: Lawlor, Elizabeth  
TITLE OF INVENTION: No. 6410286e1 tRNA Synthetase  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/018,824  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/785,076  
FILING DATE: 17-JAN-1997  
APPLICATION NUMBER: 9601096.2  
FILING DATE: 19-JAN-1996  
APPLICATION NUMBER: 9615845.6  
FILING DATE: 27-JUL-1996  
APPLICATION NUMBER: 9622617.0  
FILING DATE: 30-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimmil, Edward R  
REGISTRATION NUMBER: 38,891

REFERENCE/DOCKET NUMBER: P31354-3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 432 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-018-824-2

Query Match 42.9%; Score 39; DB 4; Length 432;  
Best Local Similarity 57.1%; Pred. No. 1.1e+02;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SPOQETVLDGNLI 15  
Db 335 NPENETVLCADLI 348

Search completed: October 10, 2003, 17:26:06  
Job time : 31 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 10, 2003, 17:24:49 ; Search time 67 Seconds  
(without alignments)  
43.288 Million cell updates/sec

Title: US-09-991-795-1

Perfect score: 91  
Sequence: 1 KSPDQOETVDGNLIRY 18

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 600653 seqs, 161128416 residues

Total number of hits satisfying chosen parameters: 600653

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubppaa/PTCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubppaa/PTCT\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubppaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubppaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubppaa/US09C\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubppaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91	100.0	18	11	US-09-992-067-1
2	79	86.8	16	10	US-09-826-290-44
3	46	50.5	601	15	US-10-156-761-14851
4	43	47.3	432	15	US-10-156-761-14180
5	43	47.3	946	9	US-09-828-423-3
6	42	46.2	544	9	US-09-841-132-400
7	41	45.1	444	12	US-10-032-585-7704
8	41	45.1	612	9	US-09-815-242-10107
9	41	45.1	901	11	US-09-828-062-8
10	40.5	44.5	436	9	US-09-240-675-2
11	40.5	44.5	557	9	US-09-240-675-4
12	40.5	44.5	575	10	US-09-925-300-1672
13	40	44.0	117	15	US-10-106-698-5879
14	40	44.0	131	12	US-10-097-111-330
15	40	44.0	241	10	US-09-738-626-6844

16	40	44.0	486	15	US-10-156-761-9284	Sequence 9284, Ap
17	40	44.0	520	12	US-10-167-831-1	Sequence 1, Appl1
18	40	44.0	654	15	US-10-043-344-12	Sequence 12, Appl1
19	40	44.0	769	10	US-09-818-247-3	Sequence 3, Appl1
20	40	44.0	769	10	US-09-982-107-10	Sequence 10, Appl1
21	40	44.0	769	12	US-09-949-039-69	Sequence 69, Appl1
22	39.5	43.4	72	14	US-10-032-159A-4	Sequence 4, Appl1
23	39.5	43.4	366	14	US-10-032-159A-2	Sequence 2, Appl1
24	39.5	43.4	366	11	US-10-032-159A-20	Sequence 20, Appl1
25	39.5	43.4	536	11	US-09-798-412-2	Sequence 2, Appl1
26	39.5	43.4	536	11	US-09-798-412-5	Sequence 5, Appl1
27	39.5	43.4	536	15	US-10-325-917-2	Sequence 2, Appl1
28	39.5	43.4	536	15	US-10-325-917-5	Sequence 5, Appl1
29	39	42.9	378	15	US-10-156-761-11744	Sequence 11744, A
30	39	42.9	516	12	US-10-032-201B-289	Sequence 289, App
31	39	42.9	604	12	US-10-171-404A-6	Sequence 6, Appl1
32	38.5	42.3	323	15	US-10-153-668-240	Sequence 240, App
33	38.5	42.3	352	15	US-10-077-584-8	Sequence 463, App
34	38	41.8	60	9	US-09-860-332A-12	Sequence 12, Appl1
35	38	41.8	133	10	US-09-731-872-463	Sequence 463, App
36	38	41.8	133	12	US-09-876-997-463	Sequence 463, App
37	38	41.8	156	15	US-10-101-464A-522	Sequence 522, App
38	38	41.8	230	9	US-09-815-242-5264	Sequence 5264, App
39	38	41.8	238	9	US-09-815-242-12239	Sequence 12239, A
40	38	41.8	286	9	US-09-764-898-256	Sequence 256, App
41	38	41.8	301	12	US-10-148-907A-21	Sequence 21, Appl1
42	38	41.8	348	11	US-09-796-753-34	Sequence 34, Appl1
43	38	41.8	452	12	US-10-301-822-79	Sequence 79, Appl1
44	38	41.8	453	11	US-09-796-753-32	Sequence 32, Appl1
45	38	41.8	502	11	US-09-931-836-40	Sequence 40, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-992-067-1  
; Sequence 1, Application US/09992067  
; Publication No. US20030096431A1  
; GENERAL INFORMATION:  
; APPLICANT: Jackowski, George  
; TITLE OF INVENTION: Interleukin Alpha Trypsin Inhibitor Biopolymer Markers Indicative of Insi  
; TITLE OF INVENTION: Resistance  
; FILE REFERENCE: 2132.100  
; CURRENT APPLICATION NUMBER: US/09/992.067  
; CURRENT FILING DATE: 2001-11-21  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-992-067-1

Query Match 100.0%; Score 91; DB 11; Length 18;  
Best Local Similarity 100.0%; Pred. No. 8,4e-09;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KSPDQOETVDGNLIRY 18  
Db 1 KSPDQOETVDGNLIRY 18  
RESULT 2  
US-09-826-290-44  
; Sequence 44, Application US/09826290  
; Patent No. US2002016468A1  
; GENERAL INFORMATION:  
; APPLICANT: Durham, L. Kathryn  
; APPLICANT: Friedman, David L.  
; APPLICANT: Herath, Herath Mudiyanselage Athula Chandrasiri  
; APPLICANT: Kimmel, Lida H.  
; APPLICANT: Parekh, Rajesh Bhikhu

```
; APPLICANT: Potter, David M.
; APPLICANT: Rohlf, Christian
; APPLICANT: Silber, B. Michael
; APPLICANT: Stiger, Thomas R.
; APPLICANT: Sunderland, P. Trey
; APPLICANT: Townsend, Robert Reid
; APPLICANT: White, Frost
; APPLICANT: Williams, Stephen A.
; TITLE OF INVENTION: Nucleic Acid Molecules, Polypeptides and
; TITLE OF INVENTION: Uses Thereof, Including Diagnosis and Treatment of
; FILE REFERENCE: 2572-1-001 N2
; CURRENT APPLICATION NUMBER: US/09/826,290
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: US 60/194,504
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: US 60/253,647
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 492
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 16
; TYPE: PRT
; ORGANISM: homo sapien
US-09-826-290-44
```

```
Query Match      86.8%; Score 79; DB 10; Length 16;
Best Local Similarity 100.0%; Pred. No. 8e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      2 SPEGQETVLDDGNLIIR 17
Db      1 SPEGQETVLDDGNLIIR 16
```

```
RESULT 3
US-10-156-761-14851
; Sequence 14851, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14851
; LENGTH: 601
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14851
```

```
Query Match      50.5%; Score 46; DB 15; Length 601;
Best Local Similarity 53.3%; Pred. No. 22;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```
Oy      3 PEGQETVLDDGNLIIR 17
Db      57 PEGQETVLDDGNLIIR 71
```

```
RESULT 4
US-10-156-761-14180
```

```
; Sequence 14180, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14180
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14180
```

```
Query Match      47.3%; Score 43; DB 15; Length 432;
Best Local Similarity 37.5%; Pred. No. 48;
Matches 6; Conservative 8; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy      2 SPEGQETVLDDGNLIIR 17
Db      397 APVDVDTVDDGNLIIR 412
```

```
RESULT 5
US-09-828-423-3
; Sequence 3, Application US/09828423
; Patent No. US20020099178A1
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; Guegler, Karl J.
; Patterson, Chandra
; TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE
; INHIBITOR HEAVY CHAIN PRECURSOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESS: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/828,423
; FILING DATE: 05-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/388,774
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Certone, Michael C
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0505 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 3:
```

;  
; PRIOR APPLICATION NUMBER: 60/  
; PRIOR FILING DATE: 2000-04-07

NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO: 8  
LENGTH: 901  
TYPE: PRT  
ORGANISM: Physcomitrella patens  
US-09-828-062-8

Query Match 45.1%; Score 41; DB 11; Length 901;  
Best Local Similarity 43.8%; Pred. No. 2.5e+02;  
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 KSPQKVEVDIIDNFIIRW 16  
Db 580 KDPETRETVLESALV 595

RESULT 10  
US-09-240-675-2  
Sequence 2, Application US/09240675  
Patent No. US20020055492A1  
GENERAL INFORMATION:  
APPLICANT: BENOIT, Patrick  
APPLICANT: MEYER, Francois  
APPLICANT: MAGUIRE, Deborah  
APPLICANT: TOVEY, Michael G.  
TITLE OF INVENTION: MONOCLONAL ANTIBODIES AGAINST THE INTERFERON  
RECEPTOR, WITH NEUTRALIZING ACTIVITY AGAINST TYPE I  
TITLE OF INVENTION: INTERFERON  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
ZIP: 20007  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/240.675  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/307.588  
FILING DATE: 05-DEC-1994  
APPLICATION NUMBER: PCT/EP93/00770  
FILING DATE: 30-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 92400902.0  
FILING DATE: 31-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: SAXE, Bernhard D.  
REGISTRATION NUMBER: 28,665  
REFERENCE/DOCKET NUMBER: 17283/117/GUPL  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 436 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-240-675-2

Query Match 44.5%; Score 40.5; DB 9; Length 436;  
Best Local Similarity 42.1%; Pred. No. 1.3e+02;  
Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

QY 1 KSPQKVEVDIIDNFIIRW 18

Db 31 KSPQKVEVDIIDNFIIRW 49

RESULT 11  
US-09-240-675-4  
Sequence 4, Application US/09240675  
Patent No. US20020055492A1  
GENERAL INFORMATION:  
APPLICANT: BENOIT, Patrick  
APPLICANT: MEYER, Francois  
APPLICANT: MAGUIRE, Deborah  
APPLICANT: TOVEY, Michael G.  
TITLE OF INVENTION: MONOCLONAL ANTIBODIES AGAINST THE INTERFERON  
RECEPTOR, WITH NEUTRALIZING ACTIVITY AGAINST TYPE I  
TITLE OF INVENTION: INTERFERON  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
ZIP: 20007  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/240.675  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/307.588  
FILING DATE: 05-DEC-1994  
APPLICATION NUMBER: PCT/EP93/00770  
FILING DATE: 30-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 92400902.0  
FILING DATE: 31-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: SAXE, Bernhard D.  
REGISTRATION NUMBER: 28,665  
REFERENCE/DOCKET NUMBER: 17283/117/GUPL  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 557 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-240-675-4

Query Match 44.5%; Score 40.5; DB 9; Length 557;  
Best Local Similarity 42.1%; Pred. No. 1.7e+02;  
Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

QY 1 KSPQKVEVDIIDNFIIRW 18  
Db 31 KSPQKVEVDIIDNFIIRW 49

RESULT 12  
US-09-925-300-1672  
Sequence 1672, Application US/09925300  
Patent No. US20020151681A1  
GENERAL INFORMATION:  
APPLICANT: Craig Rosen,  
APPLICANT: Steve Ruden,  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA101

;; CURRENT APPLICATION NUMBER: US/09/925,300  
;; CURRENT FILING DATE: 2001-08-10  
;; PRIOR APPLICATION NUMBER: PCT/US00/05988  
;; PRIOR FILING DATE: 2000-03-08  
;; PRIOR APPLICATION NUMBER: 60/124,270  
;; PRIOR FILING DATE: 1999-03-12  
;; NUMBER OF SEQ ID NOS: 1890  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO: 1672  
;; LENGTH: 575  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: SITE  
;; LOCATION: (186)  
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
;; NAME/KEY: SITE  
;; LOCATION: (555)  
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-300-1672

Query Match 44.5%; Score 40.5; DB 10; Length 575;  
Best Local Similarity 42.1%; Pred. No. 1.8e+02;  
Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

QY 1 KSPQOETVLDGNLIRY 18  
Db 49 KSPQKEVDIDDFILRW 67

RESULT 13  
US-10-106-698-5879  
;; Sequence 5879, Application US/10106698  
;; Publication No. US20030109690A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Ruben et al.  
;; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide  
;; FILE REFERENCE: PA005P1  
;; CURRENT APPLICATION NUMBER: US/10/106,698  
;; CURRENT FILING DATE: 2002-03-27  
;; PRIOR APPLICATION NUMBER: PCT/US00/26524  
;; PRIOR FILING DATE: 2000-09-28  
;; PRIOR APPLICATION NUMBER: US 60/157,137  
;; PRIOR FILING DATE: 1999-09-29  
;; PRIOR APPLICATION NUMBER: US 60/163,280  
;; PRIOR FILING DATE: 1999-11-03  
;; NUMBER OF SEQ ID NOS: 8564  
;; SOFTWARE: PatentIn Ver. 3.0  
;; SEQ ID NO 5879  
;; LENGTH: 117  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-106-698-5879

Query Match 44.0%; Score 40; DB 15; Length 117;  
Best Local Similarity 50.0%; Pred. No. 34;  
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 PEOQETVLDGNLII 16  
Db 87 PESQEGLLPGLVY 100

RESULT 14  
US-10-097-111-330  
;; Sequence 330, Application US/10097111  
;; Publication No. US20030138771A1  
;; GENERAL INFORMATION:  
;; APPLICANT: PELLETIER, JERRY  
;; APPLICANT: GROS, PHILIPPE  
;; APPLICANT: DUBOW, MICHAEL  
;; TITLE OF INVENTION: DNA SEQUENCES FROM S. PNEUMONIAE BACTERIOPHAGE EPI THAT  
;; TITLE OF INVENTION: ENCODE ANTI-MICROBIAL POLYPEPTIDES

;; FILE REFERENCE: 073406-0603  
;; CURRENT APPLICATION NUMBER: US/10/097,111  
;; CURRENT FILING DATE: 2002-07-24  
;; PRIOR APPLICATION NUMBER: 09/676,412  
;; PRIOR FILING DATE: 2000-09-29  
;; PRIOR APPLICATION NUMBER: 60/157,218  
;; PRIOR FILING DATE: 1999-09-30  
;; NUMBER OF SEQ ID NOS: 552  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 330  
;; LENGTH: 131  
;; TYPE: PRT  
;; ORGANISM: Streptococcus pneumoniae  
US-10-097-111-330

Query Match 44.0%; Score 40; DB 12; Length 131;  
Best Local Similarity 52.9%; Pred. No. 39;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 KSPQOETVLDGNLIR 17  
Db 94 KAVTLEIILIDGNIR 110

RESULT 15  
US-09-738-626-6844  
;; Sequence 6844, Application US/09738626  
;; Publication No. US20020197605A1  
;; GENERAL INFORMATION:  
;; APPLICANT: NAKAGAWA, SATOSHI  
;; APPLICANT: MIZOGUCHI, HIROSHI  
;; APPLICANT: ANDO, SEIKO  
;; APPLICANT: HAYASHI, MIKIRO  
;; APPLICANT: OCHIAI, KEIKO  
;; APPLICANT: YOKOI, HARUHIKO  
;; APPLICANT: TATEISHI, NAOKO  
;; APPLICANT: SENOH, AKIHIRO  
;; APPLICANT: IKEDA, MASATO  
;; APPLICANT: OZAKI, AKIO  
;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
;; FILE REFERENCE: 249-125  
;; CURRENT APPLICATION NUMBER: US/09/738,626  
;; CURRENT FILING DATE: 2000-12-18  
;; PRIOR APPLICATION NUMBER: JP 99/377484  
;; PRIOR FILING DATE: 1999-12-16  
;; PRIOR APPLICATION NUMBER: JP 00/159162  
;; PRIOR FILING DATE: 2000-04-07  
;; PRIOR APPLICATION NUMBER: JP 00/280988  
;; PRIOR FILING DATE: 2000-08-03  
;; NUMBER OF SEQ ID NOS: 7059  
;; SOFTWARE: PatentIn Ver. 3.0  
;; SEQ ID NO 6844  
;; LENGTH: 241  
;; TYPE: PRT  
;; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-6844

Query Match 44.0%; Score 40; DB 10; Length 241;  
Best Local Similarity 47.4%; Pred. No. 79;  
Matches 9; Conservative 4; Mismatches 4; Indels 2; Gaps 1;

QY 2 SPOQETVLDGNL--IIRY 18  
Db 200 APQDEVVVGSLAGIVRY 218

RESULT 16  
US-10-156-761-9284  
;; Sequence 9284, Application US/10156761  
;; Publication No. US20030119018A1  
;; GENERAL INFORMATION:  
;; APPLICANT: OMURA, SATOSHI  
;; APPLICANT: IKEDA, HARUO

```
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9284
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-9284
```

```
Query Match          44.0%; Score 40; DB 15; Length 486;
Best Local Similarity 37.5%; Pred. No. 1.8e+02;
Matches 6; Conservative 7; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 KSPSEQETVLDGNLI 16
Db      169 RTPEGERVLEADLVV 184
```

```
RESULT 17
US-10-167-831-1
; Sequence 1, Application US/10167831
; Publication No. US20030166209A1
; GENERAL INFORMATION:
; APPLICANT: Shuman, Stewart
; APPLICANT: Ho, C. Kiong
; TITLE OF INVENTION: Pharmacological Targeting of mRNA Cap Formation
; FILE REFERENCE: D6388D
; CURRENT APPLICATION NUMBER: US/10/167,831
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 09/752,165
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 104
; SEQ ID NO 1
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: PEPTIDE
; OTHER INFORMATION: Plasmodium falciparum RNA guanylyltransferase Pgt1
; US-10-167-831-1
```

```
Query Match          44.0%; Score 40; DB 12; Length 520;
Best Local Similarity 57.1%; Pred. No. 1.9e+02;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      2 SPEQETVLDGNLI 15
Db      100 SKKQQLTLDGELV 113
```

```
RESULT 18
US-10-043-344-12
; Sequence 12, Application US/10043344
; Publication No. US20030088086A1
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Harkness, Robin E.
; APPLICANT: Schryvers, Anthony B.
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Mordin, Andrew D.
```

```
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: TRANSFERRIN RECEPTOR GENES
; FILE REFERENCE: 1038-1221 MIS
; CURRENT APPLICATION NUMBER: US/10/043,344
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: 08/649,518
; PRIOR FILING DATE: 1996-05-17
; NUMBER OF SEQ ID NOS: 160
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 654
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; US-10-043-344-12
```

```
Query Match          44.0%; Score 40; DB 15; Length 654;
Best Local Similarity 47.4%; Pred. No. 2.5e+02;
Matches 9; Conservative 4; Mismatches 2; Indels 4; Gaps 1;
```

```
QY      1 KSPSEQ-----ETVLDGNLI 15
Db      330 QNPENQKLTSTETLIDGKLI 348
```

```
RESULT 19
US-09-818-247-3
; Sequence 3, Application US/09818247
; Patent No. US20020102657A1
; GENERAL INFORMATION:
; APPLICANT: Mostlov, Keith E.
; APPLICANT: Chaplin, Steven J.
; APPLICANT: Richman-Eisenstat, Janice
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Ligands directed to the No. US20020102657A1-Stralk Region of p19R and Methods of Use Ther
; FILE REFERENCE: 18062E-000910US
; CURRENT APPLICATION NUMBER: US/09/818,247
; CURRENT FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: WO PCT/US01/09699
; PRIOR FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: US 60/192,197
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: US 60/192,198
; PRIOR FILING DATE: 2000-03-27
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 769
; TYPE: PRT
; ORGANISM: Rattus sp.
; FEATURE:
; OTHER INFORMATION: rat polymeric immunoglobulin receptor (p19R)
; US-09-818-247-3
```

```
Query Match          44.0%; Score 40; DB 10; Length 769;
Best Local Similarity 63.6%; Pred. No. 3e+02;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      2 SPEQETVLDG 12
Db      706 TPDQETVLEG 716
```

```
RESULT 20
US-09-982-107-10
; Sequence 10, Application US/09982107
; Patent No. US20020159958A1
; GENERAL INFORMATION:
; APPLICANT: HIATT, ANDREW C.
; APPLICANT: HEIN, MIC B.
; TITLE OF INVENTION: METHODS FOR PRODUCING IMMUNOGLOBULINS CONTAINING
; TITLE OF INVENTION: PROTECTION PROTEINS IN PLANTS AND THEIR USE
; FILE REFERENCE: EPI3002E
```



CURRENT APPLICATION NUMBER: US/09/982,107  
 CURRENT FILING DATE: 2001-10-16  
 NUMBER OF SEQ ID NOS: 19  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 10  
 LENGTH: 769  
 TYPE: PRT  
 ORGANISM: Rattus sp.  
 US-09-982-107-10

Query Match 44.0%; Score 40; DB 10; Length 769;  
 Best Local Similarity 63.6%; Pred. No. 3e+02; 1; Indels 0; Gaps 0;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 SPEQOETVLDPG 12  
 Db 706 TPDTQETVLEGS 716

RESULT 21  
 US-09-949-039-69  
 Sequence 69, Application US/09949039  
 Publication No. US20030166160A1  
 GENERAL INFORMATION:  
 APPLICANT: HAWLEY, STEPHEN B.  
 TITLE OF INVENTION: COMPOUNDS AND MOLECULAR COMPLEXES COMPRISING MULTIPLE  
 TITLE OF INVENTION: BINDING REGIONS DIRECTED TO TRANSCYTOTIC LIGANDS  
 FILE REFERENCE: 057220/1301  
 CURRENT APPLICATION NUMBER: US/09/949,039  
 CURRENT FILING DATE: 2001-09-06  
 NUMBER OF SEQ ID NOS: 114  
 SOFTWARE: PatentIn Ver. 2.1  
 SEQ ID NO 69  
 LENGTH: 769  
 TYPE: PRT  
 ORGANISM: Rattus sp.  
 US-09-949-039-69

Query Match 44.0%; Score 40; DB 12; Length 769;  
 Best Local Similarity 63.6%; Pred. No. 3e+02; 1; Indels 0; Gaps 0;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 SPEQOETVLDPG 12  
 Db 706 TPDTQETVLEGS 716

RESULT 22  
 US-10-032-159A-4  
 Sequence 4, Application US/10032159A  
 Publication No. US20020164703A1  
 GENERAL INFORMATION:  
 APPLICANT: Pawlowski, Krzysztof  
 APPLICANT: Reed, John C.  
 APPLICANT: Godzik, Adam  
 TITLE OF INVENTION: CARD-DOMAIN CONTAINING POLYPEPTIDES,  
 TITLE OF INVENTION: ENCODING NUCLEIC ACIDS, AND METHODS OF USE  
 FILE REFERENCE: P-LJ 5100  
 CURRENT APPLICATION NUMBER: US/10/032,159A  
 CURRENT FILING DATE: 2001-12-19  
 PRIOR APPLICATION NUMBER: US 60/257,457  
 PRIOR FILING DATE: 2000-12-21  
 NUMBER OF SEQ ID NOS: 37  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 4  
 LENGTH: 72  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-032-159A-4

Query Match 43.4%; Score 39.5; DB 14; Length 72;  
 Best Local Similarity 52.9%; Pred. No. 24;  
 Matches 9; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 2 SPEQOETVL-DGNLIIR 17  
 Db 30 NPDDDEQVLSDPNLVIR 46

RESULT 23  
 US-10-032-159A-2  
 Sequence 2, Application US/10032159A  
 Publication No. US20020164703A1  
 GENERAL INFORMATION:  
 APPLICANT: Pawlowski, Krzysztof  
 APPLICANT: Reed, John C.  
 APPLICANT: Godzik, Adam  
 TITLE OF INVENTION: CARD-DOMAIN CONTAINING POLYPEPTIDES,  
 TITLE OF INVENTION: ENCODING NUCLEIC ACIDS, AND METHODS OF USE  
 FILE REFERENCE: P-LJ 5100  
 CURRENT APPLICATION NUMBER: US/10/032,159A  
 CURRENT FILING DATE: 2001-12-19  
 PRIOR APPLICATION NUMBER: US 60/257,457  
 PRIOR FILING DATE: 2000-12-21  
 NUMBER OF SEQ ID NOS: 37  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 2  
 LENGTH: 366  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-032-159A-2

Query Match 43.4%; Score 39.5; DB 14; Length 366;  
 Best Local Similarity 52.9%; Pred. No. 1.6e+02;  
 Matches 9; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 2 SPEQOETVL-DGNLIIR 17  
 Db 41 NPDDDEQVLSDPNLVIR 57

RESULT 24  
 US-10-032-159A-20  
 Sequence 20, Application US/10032159A  
 Publication No. US20020164703A1  
 GENERAL INFORMATION:  
 APPLICANT: Pawlowski, Krzysztof  
 APPLICANT: Reed, John C.  
 APPLICANT: Godzik, Adam  
 TITLE OF INVENTION: CARD-DOMAIN CONTAINING POLYPEPTIDES,  
 TITLE OF INVENTION: ENCODING NUCLEIC ACIDS, AND METHODS OF USE  
 FILE REFERENCE: P-LJ 5100  
 CURRENT APPLICATION NUMBER: US/10/032,159A  
 CURRENT FILING DATE: 2001-12-19  
 PRIOR APPLICATION NUMBER: US 60/257,457  
 PRIOR FILING DATE: 2000-12-21  
 NUMBER OF SEQ ID NOS: 37  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 20  
 LENGTH: 366  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-10-032-159A-20

Query Match 43.4%; Score 39.5; DB 14; Length 366;  
 Best Local Similarity 52.9%; Pred. No. 1.6e+02;  
 Matches 9; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 2 SPEQOETVL-DGNLIIR 17  
 Db 41 NPDDDEQVLSDPNLVIR 57

RESULT 25  
 US-09-798-412-2  
 Sequence 2, Application US/09798412

```
Publication No. US20030109428A1
GENERAL INFORMATION:
APPLICANT: Bertin, John
TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
FILE REFERENCE: 0734-327001
CURRENT APPLICATION NUMBER: US/09/798,412
PRIOR FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: US 09/728,260
PRIOR FILING DATE: 2000-12-01
PRIOR APPLICATION NUMBER: US 09/685,791
PRIOR FILING DATE: 2000-10-10
PRIOR APPLICATION NUMBER: US 09/513,904
PRIOR FILING DATE: 2000-02-25
PRIOR APPLICATION NUMBER: US 09/507,533
PRIOR FILING DATE: 2000-02-18
PRIOR APPLICATION NUMBER: US 60/168,780
PRIOR FILING DATE: 1999-12-03
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 536
TYPE: PRT
ORGANISM: Rattus norvegicus
US-09-798-412-2
```

```
Query Match 43.4%; Score 39.5; DB 11; Length 536;
Best Local Similarity 52.9%; Pred. No. 2.4e+02;
Matches 9; Conservative 4; Mismatches 3; Indels 1; Gaps 1;
```

```
Oy 2 SPOQETVL-DGNLIIR 17
Db 41 NPDDSEQVLSDPNIVIR 57
```

```
RESULT 26
US-09-798-412-5
Sequence 5, Application US/09798412
Publication No. US20030109428A1
GENERAL INFORMATION:
APPLICANT: Bertin, John
TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
FILE REFERENCE: 0734-327001
CURRENT APPLICATION NUMBER: US/09/798,412
PRIOR FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: US 09/728,260
PRIOR FILING DATE: 2000-12-01
PRIOR APPLICATION NUMBER: US 09/685,791
PRIOR FILING DATE: 2000-10-10
PRIOR APPLICATION NUMBER: US 09/513,904
PRIOR FILING DATE: 2000-02-25
PRIOR APPLICATION NUMBER: US 09/507,533
PRIOR FILING DATE: 2000-02-18
PRIOR APPLICATION NUMBER: US 60/168,780
PRIOR FILING DATE: 1999-12-03
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 536
TYPE: PRT
ORGANISM: Homo sapiens
US-09-798-412-5
```

```
Query Match 43.4%; Score 39.5; DB 11; Length 536;
Best Local Similarity 52.9%; Pred. No. 2.4e+02;
Matches 9; Conservative 4; Mismatches 3; Indels 1; Gaps 1;
```

```
Oy 2 SPOQETVL-DGNLIIR 17
Db 41 NPDDSEQVLSDPNIVIR 57
```

```
RESULT 27
US-10-325-917-2
Sequence 2, Application US/10325917
Publication No. US20030113787A1
GENERAL INFORMATION:
APPLICANT: Bertin, John
TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
FILE REFERENCE: 0734-327001
CURRENT APPLICATION NUMBER: US/10/325,917
PRIOR FILING DATE: 2002-12-20
PRIOR APPLICATION NUMBER: US/09/798,412
PRIOR FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: US 09/728,260
PRIOR FILING DATE: 2000-12-01
PRIOR APPLICATION NUMBER: US 09/685,791
PRIOR FILING DATE: 2000-10-10
PRIOR APPLICATION NUMBER: US 09/513,904
PRIOR FILING DATE: 2000-02-25
PRIOR APPLICATION NUMBER: US 09/507,533
PRIOR FILING DATE: 2000-02-18
PRIOR APPLICATION NUMBER: US 60/168,780
PRIOR FILING DATE: 1999-12-03
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 536
TYPE: PRT
ORGANISM: Rattus norvegicus
US-10-325-917-2
```

```
Query Match 43.4%; Score 39.5; DB 15; Length 536;
Best Local Similarity 52.9%; Pred. No. 2.4e+02;
Matches 9; Conservative 4; Mismatches 3; Indels 1; Gaps 1;
```

```
Oy 2 SPOQETVL-DGNLIIR 17
Db 41 NPDDSEQVLSDPNIVIR 57
```

```
RESULT 28
US-10-325-917-5
Sequence 5, Application US/10325917
Publication No. US20030113787A1
GENERAL INFORMATION:
APPLICANT: Bertin, John
TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
FILE REFERENCE: 0734-327001
CURRENT APPLICATION NUMBER: US/10/325,917
PRIOR FILING DATE: 2002-12-20
PRIOR APPLICATION NUMBER: US/09/798,412
PRIOR FILING DATE: 2001-03-02
PRIOR APPLICATION NUMBER: US 09/728,260
PRIOR FILING DATE: 2000-12-01
PRIOR APPLICATION NUMBER: US 09/685,791
PRIOR FILING DATE: 2000-10-10
PRIOR APPLICATION NUMBER: US 09/513,904
PRIOR FILING DATE: 2000-02-25
PRIOR APPLICATION NUMBER: US 09/507,533
PRIOR FILING DATE: 2000-02-18
PRIOR APPLICATION NUMBER: US 60/168,780
PRIOR FILING DATE: 1999-12-03
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 536
TYPE: PRT
ORGANISM: Homo sapiens
US-10-325-917-5
```

```
Query Match 43.4%; Score 39.5; DB 15; Length 536;
Best Local Similarity 52.9%; Pred. No. 2.4e+02;
```

Matches 9; Conservative 4; Mismatches 3; Indels 1; Gaps 1;  
Qy 2 SPEQOETVL-DGNLIIR 17  
Db 41 NPDEEQVLSDPNLVIR 57

RESULT 29  
US-10-156-761-11744  
; Sequence 11744, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 11744  
; LENGTH: 378  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-11744

Query Match 42.9%; Score 39; DB 15; Length 378;  
Best Local Similarity 42.1%; Pred. No. 2e+02;  
Matches 8; Conservative 6; Mismatches 1; Indels 4; Gaps 1;  
Qy 2 SPEQOETVL---DGNLIIR 16  
Db 93 SPEQERLRPEIGDSVV 111

RESULT 30  
US-10-032-201B-289  
; Sequence 289, Application US/10032201B  
; Publication No. US20030167524A1  
; GENERAL INFORMATION:  
; APPLICANT: Van Rooijen, Gijls  
; APPLICANT: Deckers, Harm  
; APPLICANT: Heifetz, Peter Bernard  
; APPLICANT: Briggs, Steven  
; APPLICANT: Dalmia, Bipin Kumar  
; APPLICANT: Del Val, Greg  
; APPLICANT: Zaplachinski, Steve  
; APPLICANT: Moloney, Maurice  
; TITLE OF INVENTION: METHODS FOR THE PRODUCTION OF MULTIMERIC PROTEINS, AND RELATED  
; FILE REFERENCE: 3814 351B  
; CURRENT APPLICATION NUMBER: US/10/032,201B  
; CURRENT FILING DATE: 2001-12-19  
; NUMBER OF SEQ ID NOS: 313  
; SOFTWARE: Patscseq for Windows Version 4.0  
; SEQ ID NO 289  
; LENGTH: 516  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-10-032-201B-289

Query Match 42.9%; Score 39; DB 12; Length 516;  
Best Local Similarity 43.8%; Pred. No. 2.8e+02;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 3 PEQOETVLDGNLIIR 18  
Db 275 PKAVERQADGRLVIR 290

RESULT 31  
US-10-171-404A-6  
; Sequence 6, Application US/10171404A  
; Publication No. US20030177529A1  
; GENERAL INFORMATION:  
; APPLICANT: BASF PLANT SCIENCE GMBH  
; TITLE OF INVENTION: SUGAR AND LIPID METABOLISM REGULATORS IN PLANTS II  
; FILE REFERENCE: 16313-0119  
; CURRENT APPLICATION NUMBER: US/10/171,404A  
; CURRENT FILING DATE: 2002-11-19  
; PRIOR APPLICATION NUMBER: 60/295,680  
; PRIOR FILING DATE: 2001-06-04  
; NUMBER OF SEQ ID NOS: 99  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 604  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-10-171-404A-6

Query Match 42.9%; Score 39; DB 12; Length 604;  
Best Local Similarity 61.5%; Pred. No. 3.4e+02;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 KSPEQOETVLDGN 13  
Db 279 KPKRKQKETSJDGN 291

RESULT 32  
US-10-153-668-240  
; Sequence 240, Application US/10153668  
; Publication No. US20030092616A1  
; GENERAL INFORMATION:  
; APPLICANT: HONDA, Goichi  
; APPLICANT: MATSUDA, Akio  
; APPLICANT: MURAMATSU, Shuji  
; APPLICANT: ISHIZAWA, Kenya  
; TITLE OF INVENTION: STAT6 Activating Gene  
; FILE REFERENCE: 1254-0207P  
; CURRENT APPLICATION NUMBER: US/10/153,668  
; CURRENT FILING DATE: 2002-05-24  
; PRIOR APPLICATION NUMBER: US 60/293,172  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: US 60/316,031  
; PRIOR FILING DATE: 2001-08-31  
; PRIOR APPLICATION NUMBER: US 60/328,403  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: JP 2001-157043  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: JP 2001-260681  
; PRIOR FILING DATE: 2001-08-30  
; PRIOR APPLICATION NUMBER: JP 2001-313175  
; NUMBER OF SEQ ID NOS: 488  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 240  
; LENGTH: 323  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-153-668-240

Query Match 42.3%; Score 38.5; DB 15; Length 323;  
Best Local Similarity 47.1%; Pred. No. 2e+02;  
Matches 8; Conservative 5; Mismatches 3; Indels 1; Gaps 1;  
Qy 2 SPEQOETVLDGNLIIR 18  
Db 275 PKAVERQADGRLVIR 290

```
Db      238 TPBOO-GILEKELVRY 253

RESULT 33
US-10-077-584-8
; Sequence 8, Application US/10077584
; Publication No. US20030073610A1
; GENERAL INFORMATION:
; APPLICANT: LINDQUIST, SUSAN
; APPLICANT: KROBITSCH, SYLVIA
; APPLICANT: OUTEIRO, TIAGO F.
; TITLE OF INVENTION: YEAST SCREENS FOR THE TREATMENT OF HUMAN DISEASE
; FILE REFERENCE: ARCD:367US
; CURRENT APPLICATION NUMBER: US/10/077,584
; CURRENT FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/269,157
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patencin Ver. 2.1
; SEQ ID NO 8
; LENGTH: 352
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-077-584-8

Query Match      42.3%; Score 38.5; DB 15; Length 352;
Best Local Similarity 44.4%; Pred. No. 2.2e+02;
Matches 8; Conservative 4; Mismatches 1; Indels 5; Gaps 1;

QY      1 KSPBOQETVLDGNLIR 18
        |||:|:|:|:|:|
Db      320 KNPSOR----GNLIVKY 332

RESULT 34
US-09-860-232A-12
; Sequence 12, Application US/09860232A
; Patent No. US2002028494A1
; GENERAL INFORMATION:
; APPLICANT: CURTIS, ROY A. J.
; TITLE OF INVENTION: 57256 AND 58289, NOVEL HUMAN
; FILE REFERENCE: 381552001500
; CURRENT APPLICATION NUMBER: US/09/860,232A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: US 60/205,288
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus amino acid
US-09-860-232A-12

Query Match      41.8%; Score 38; DB 9; Length 60;
Best Local Similarity 58.3%; Pred. No. 35;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      3 PEQGETVLDGNL 14
        |||:|:|:|
Db      47 PENQNFLEGNL 58

RESULT 35
US-09-731-872-463
; Sequence 463, Application US/09731872
; Patent No. US20020102604A1
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Bougueleret, Lydie

Db      238 TPBOO-GILEKELVRY 253

APPLICANT: Jobert, Severin
; TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS
; FILE REFERENCE: 78.US3.REG
; CURRENT APPLICATION NUMBER: US/09/731,872
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,629
; PRIOR FILING DATE: 1999-12-08
; PRIOR APPLICATION NUMBER: US 60/187,470
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 482
; SOFTWARE: Patent .pm
; SEQ ID NO 463
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-731-872-463

Query Match      41.8%; Score 38; DB 10; Length 133;
Best Local Similarity 72.7%; Pred. No. 87;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      7 ETVLDGNLIR 17
        |||:|:|:|
Db      113 ETALYGNLIR 123

RESULT 36
US-09-876-997-463
; Sequence 463, Application US/09876997
; Publication No. US20030152921A1
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Jobert, Severin
; TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS
; FILE REFERENCE: 78.US4.CIP
; CURRENT APPLICATION NUMBER: US/09/876,997
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 09/731,872
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/187,470
; PRIOR FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: US 60/169,629
; PRIOR FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 482
; SOFTWARE: Patent .pm
; SEQ ID NO 463
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-876-997-463

Query Match      41.8%; Score 38; DB 12; Length 133;
Best Local Similarity 72.7%; Pred. No. 87;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      7 ETVLDGNLIR 17
        |||:|:|:|
Db      113 ETALYGNLIR 123

RESULT 37
US-10-101-464A-522
; Sequence 522, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; APPLICANT: Higgins, Colleen M.
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signaling
; FILE REFERENCE: 11000.1020C2
; CURRENT APPLICATION NUMBER: US/10/101,464A
```

```

; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/167,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; PRIOR FILING DATE: 2000-01-11
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 522
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
US-10-101-464A-522

Query Match          41.8%; Score 38; DB 15; Length 156;
Best Local Similarity 53.8%; Pred. No. 1.1e+02;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      2 SPEQGETVLDGNL 14
Db      21 APBSSTLVNMGNL 33

RESULT 38
US-09-815-242-5264
; Sequence 5264, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyckind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OR INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5264
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5264

Query Match          41.8%; Score 38; DB 9; Length 230;
Best Local Similarity 40.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
```

```

RESULT 39
US-09-815-242-12239
; Sequence 12239, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyckind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OR INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12239
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12239

Query Match          41.8%; Score 38; DB 9; Length 238;
Best Local Similarity 40.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY      1 KSPEQGETVLDGNLI 15
Db      146 RQPEQETVYKGSV 160

RESULT 40
US-09-764-898-256
; Sequence 256, Application US/09764898
; Patent No. US20020090673A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OR INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJZ01
; CURRENT APPLICATION NUMBER: US/09/764,898
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 311
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 256
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (32)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
```

LOCATION: (35)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (42)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (104)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-898-256

Query Match 41.8%; Score 38; DB 9; Length 286;  
Best Local Similarity 37.5%; Pred. No. 2.1e+02;  
Matches 6; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

Oy 2 SPEQOETVLDGNLIR 17  
|||:|:|:|:  
Db 228 SPSORGIGDGDILVK 243

Search completed: October 10, 2003, 17:34:46  
Job time : 68 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 10, 2003, 17:19:49 ; Search time 40 Seconds

(without alignments)  
43.276 Million cell updates/sec

Title: US-09-991-795-1

Perfect score: 91

Sequence: 1 KSPBOQETVDGNLIIRY 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91	100.0	930	2 JX0368	inter-alpha-trypsi
2	86	94.5	921	2 JX0465	inter-alpha-trypsi
3	75	82.4	45	2 S66648	major acute phase
4	57	62.6	932	2 JCS953	inter-alpha-inhibi
5	47.5	52.2	853	1 IJBONC	neural cell adhesi
6	47.5	52.2	858	1 IJRTNC	neural cell adhesi
7	45	49.5	452	2 T12981	hypothetical prote
8	44	48.4	409	2 S25068	S-locus-specific g
9	44	48.4	431	2 T14415	S-locus-specific g
10	43	47.3	596	2 T43683	devs protein - Myx
11	43	47.3	596	1 IYH2	mRNA export factor
12	43	47.3	108	2 T20389	inter-alpha-trypsi
13	42	46.2	384	2 D69368	transcription repr
14	42	46.2	428	2 T14419	S-locus-specific g
15	42	46.2	544	2 S19023	chaperonin groEL -
16	42	46.2	544	2 S86507	heat shock protein
17	42	46.2	544	2 B81565	60 kDa chaperonin
18	42	46.2	858	1 JQ1677	S-receptor kinase
19	42	46.2	1286	1 S38058	hypothetical prote
20	42	46.2	1756	1 EVECAI	DNA helicase I (EC
21	42	46.2	2515	1 A41519	posterior-group pr
22	42	46.2	150	2 T37513	probable acetyltra
23	41	45.1	242	1 S74794	hypothetical prote
24	41	45.1	276	2 AF1277	hypothetical prote
25	41	45.1	362	2 S75467	3-isopropylmalate
26	41	45.1	429	2 T07809	S-receptor kinase
27	41	45.1	436	1 JC2250	S-locus-specific g
28	41	45.1	440	2 T24478	hypothetical prote

30	41	45.1	479	2 A64117	serine-type D-Ala-
31	41	45.1	544	2 A71555	probable hsp-60 -
32	41	45.1	544	2 I40731	heat shock protein
33	41	45.1	544	2 B41479	60K heat shock pro
34	41	45.1	544	2 D81709	60 kDa chaperonin
35	41	45.1	612	2 E64820	probable oligopept
36	41	45.1	612	2 D90742	hypothetical prote
37	41	45.1	612	2 G85592	hypothetical prote
38	41	45.1	641	2 JN0800	tetracycline-minoc
39	41	45.1	682	2 S30395	gene 50, protein -
40	41	45.1	702	2 AF3038	exogenous ferric s
41	41	45.1	702	2 E98247	outer membrane pro
42	41	45.1	946	2 S54354	inter-alpha-inhibi
43	41	45.1	557	2 A32694	interferon alpha/b
44	40	44.0	209	1 E71324	probable 4-methyl-
45	40	44.0	315	2 S74357	hypothetical prote

## ALIGNMENTS

### RESULT 1

JX0368 inter-alpha-trypsin inhibitor heavy chain-related protein precursor - human

N/Alternate names: IHRP; plasma glycoprotein

C/Species: Homo sapiens (man)

C/Date: 22-Apr-1995 #sequence revision 26-May-1995 #text\_change 20-Jun-2000

C/Accession: JX0368; PC2355; S68457; S78548

R/Saguchi, K.; Tobe, T.; Hashimoto, K.; Sano, Y.; Nakano, Y.; Mura, N.H.; Tomita, M.

U. Biochem. 117, 14-18, 1995

A/Title: Cloning and characterization of cDNA for inter-alpha-trypsin inhibitor family he

A/Reference number: JX0368; MUID:95293915; PMID:7775381

A/Accession: JX0368

A/Molecule type: mRNA

A/Residues: 1-930 <SAG1>

A/Cross-references: DDBJ:D38595; NID:G664887; PIDN:BAA07602.1; PID:G1483187

A/Accession: PC2355

A/Molecule type: Protein

A/Residues: 29-44;48-55;61-75;99-111;140-151;163-169;211-224;246-267;274-281;296-329;392-

A/Experimental source: liver

R/Nishimura, H.; Kakizaki, I.; Mura, T.; Sasaki, N.; Pu, P.X.; Yamashita, T.; Nagasawa, S.

FEBS Lett. 357, 207-211, 1995

A/Title: cDNA and deduced amino acid sequence of human PK-120, a plasma kallikrein-sensit

A/Reference number: S68457; MUID:95104473; PMID:7805892

A/Accession: S68457

A/Molecule type: mRNA

A/Residues: 1-84; 'I', 86-113; 'S', 115-930 <NIS>

A/Cross-references: EMBL:D38535; NID:G624879; PIDN:BAA07536.1; PID:G1402590

A/Accession: S78548

A/Molecule type: Protein

A/Residues: 29-45;171-184;211-239;274-281;301-315;429-443;488-502;690-695;697-700;703-725;

A/Experimental source: liver

C/Comment: The amino-terminal 600 residues exhibits homology with those of inter-alpha tr

C/Function: highly sensitive to plasma kallikrein

A/Superfamily: inter-alpha-trypsin inhibitor complex component II

C/Keywords: chondroitin sulfate proteoglycan; glycoprotein; serine proteinase inhibitor

F/1-28/Domain: signal sequence #status predicted STG>

F/29-687/Product: inter-alpha-trypsin inhibitor heavy chain-related protein #status predi

F/688-930/Domain: carboxyl-terminal propeptide #status predicted <CTP>

F/81,207,517,577/Binding site: carbohydrate (Aan) (covalent) #status predicted

F/666/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

F/686,702/Binding site: carbohydrate (Ser) (covalent) #status predicted

F/701/Binding site: carbohydrate (Thr) (covalent) #status predicted

Query Match 100.0%; Score 91; DB 2; Length 930;

Best Local Similarity 100.0%; Pred. No. 4,7e-07;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 224 KSPBOQETVDGNLIIRY 241

## RESULT 2

JC4625

Inter-alpha-trypsin inhibitor heavy chain-related protein precursor - pig

C/Species: Sus scrofa domestica (domestic pig)

C/Date: 10-May-1996 #sequence\_revision 19-Jul-1996 #text\_change 20-Aug-1999

C/Accession: JC4625; PC4150

R/Hashimoto, K.; Tobe, T.; Sumiya, J.; Sano, Y.; Choi-Miura, N.H.; Ozawa, A.; Yasue, H.;

J. Biochem. 119, 577-584, 1996

A/Title: Primary structure of the pig homologue of human IHRP: Inter-alpha-trypsin inhib

A/Reference number: JC4625; MUID:96271024; PMID:8830057

A/Accession: JC4625

A/Molecule type: mRNA

A/Residues: 1-921 &lt;HAS1&gt;

A/Cross-references: GB:S82800; NID:G1836013; PIDN:AAB46821.1; PID:G1836014

A/Accession: PC4150

A/Molecule type: protein

A/Residues: 28-36;695-700/710-722 &lt;HAS2&gt;

A/Experimental source: liver

C/Comment: This protein is highly sensitive to plasma kallikrein, and is involved in acc

C/Genetics:

A/Gene: IHRP

C/Superfamily: Inter-alpha-trypsin inhibitor complex component II

C/Keywords: acute phase; glycoprotein; liver

F:1-27/Domain: signal sequence #status predicted &lt;SIG&gt;

F:28-921/Product: Inter-alpha-trypsin inhibitor family heavy chain-related protein homo

F:381-387/Region: calcium binding #status predicted

F:605-716/Region: proline-rich

F:80,205,242,513,577/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 94.5%; Score 86; DB 2; Length 921;

Best Local Similarity 88.9%; Pred. No. 3.3e-06;

Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KSPQOETVLDDGNLIRY 18

Db 222 KSPQOETVLDDGNLIRY 239

## RESULT 3

S66648

Major acute phase protein - Chilean potato-tree (fragments)

C/Species: Solanum crispum (Chilean potato-tree)

C/Date: 15-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 21-Aug-1998

C/Accession: S66648

R/Gonzalez-Ramón, N.; Alava, M.A.; Sarsa, J.A.; Pineiro, M.; Escartin, A.; Garcia-Gil, A

FEBS Lett. 371, 227-230, 1995

A/Title: The major acute phase serum protein in pigs is homologous to human plasma kali

A/Reference number: S66648; MUID:96013138; PMID:7556597

A/Accession: S66648

A/Status: Preliminary

A/Molecule type: protein

A/Residues: 1-27;28-45 &lt;GON&gt;

C/Superfamily: Inter-alpha-trypsin inhibitor complex component II

Query Match

Best Local Similarity 82.4%; Score 75; DB 2; Length 45;

Matches 14; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SPSQOETVLDDGNLIRY 18

Db 28 SPSQOETVLDDGNLIRY 44

## RESULT 4

JC5953

Inter-alpha-inhibitor H4p heavy chain - rat

C/Species: Rattus norvegicus (Norway rat)

C/Date: 05-Feb-1999 #sequence\_revision 05-Feb-1999 #text\_change 21-Jul-2000

C/Accession: JC5953

R/Soury, E.; Olivier, E.; Daveau, M.; Hiron, M.; Claeysens, S.; Ristier, J.L.; Salier, J

Biochem. Biophys. Res. Commun. 243, 522-530, 1998

A/Title: The H4p heavy chain of Inter-alpha-inhibitor family largely differs in the stru

A/Reference number: JC5953; MUID:98153798; PMID:9460842

A/Accession: JC5953

A/Status: Preliminary

A/Molecule type: mRNA

A/Residues: 1-932 &lt;SOU&gt;

A/Cross-references: GB:Y11283; NID:G2292987; PIDN:CAA7215.1; PID:G2292988

C/Superfamily: Inter-alpha-trypsin inhibitor complex component II

Query Match

Best Local Similarity 62.6%; Score 57; DB 2; Length 932;

Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KSPQOETVLDDGNLIRY 18

Db 223 KSPQOETVLDDGNLIRY 240

## RESULT 5

IUBONC

Neural cell adhesion molecule short domain form precursor - bovine

N/Alternate names: NCAM-140

C/Species: Bos primigenius taurus (cattle)

C/Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 22-Jun-1999

C/Accession: A32976; A38778; B44290; S05402

R/Linkin, V.M.; Kirantsov, N.V.; Andreeva, S.G.; Moshnyakov, M.V.; Petukhova, G.V.; Rakit

FEBS Lett. 254, 69-73, 1989

A/Title: Calmodulin-independent bovine brain adenylylate cyclase. Amino acid sequence and r

A/Reference number: A32976; MUID:69378239; PMID:276887

A/Accession: A32976

A/Molecule type: mRNA

A/Residues: 1-853 &lt;LIP&gt;

A/Cross-references: GB:X16451; NID:G60; PIDN:CAA34470.1; PID:G61

A/Accession: A38778

A/Molecule type: protein

A/Residues: 20-35;51-61;113-117;122-147;155-161;262-275;279-302;353-360;369-382;544-562;5

A/Note: the authors identified this protein as calmodulin-independent adenylylate cyclase

R/Rougou, G.; Marshak, D.R.

U. Biol. Chem. 261, 3396-3401, 1986

A/Title: Structural and immunological characterization of the amino-terminal domain of me

A/Reference number: A44290; MUID:66140120; PMID:3512556

A/Accession: B44290

A/Molecule type: protein

A/Residues: 20-36 &lt;ROU&gt;

A/Note: 23-Glu was also found

C/Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mole

C/Superfamily: Various forms of NCAM are produced by alternative splicing.

C/Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding; sit

F:1-19/Domain: signal sequence #status predicted &lt;SIG&gt;

F:20-853/Product: neural cell adhesion molecule, short domain form #status experimental

F:20-719/Domain: extracellular #status predicted &lt;EXT&gt;

F:34-98/Domain: immunoglobulin homology &lt;IMM1&gt;

F:113-191/Domain: immunoglobulin homology &lt;IMM2&gt;

F:152-156/Region: heparin binding #status predicted

F:161-165/Region: heparin binding #status predicted

F:228-288/Domain: immunoglobulin homology &lt;IMM3&gt;

F:261-270/Region: NCAM binding #status predicted

F:321-396/Domain: immunoglobulin homology &lt;IMM4&gt;

F:428-480/Domain: immunoglobulin homology &lt;IMM5&gt;

F:527-604/Domain: fibronectin type III repeat homology &lt;FN3A&gt;

F:633-693/Domain: fibronectin type III repeat homology &lt;FN3B&gt;

F:720-737/Domain: transmembrane #status predicted &lt;TM&gt;

F:738-853/Domain: intracellular #status predicted &lt;INT&gt;

F:41-96;139-189;235-286;328-394;435-488/Disulfide bonds: #status predicted

F:222,314,346,432,458,487/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 52.2%; Score 47.5; DB 1; Length 853;

Best Local Similarity 60.0%; Pred. No. 11;

Matches 9; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

Qy 3 PEOQETVLDDGNLIR 17

Db 358 PEOQET-LDGNMVR 371



## RESULT 6

JURNIC

neural cell adhesion molecule short domain form precursor - rat

N/Alternate names: NCAM-140

C/Species: Rattus norvegicus (Norway rat)

C/Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 22-Jun-1999

C/Accession: S00846; B37795; I58136

R/Small, S.J.; Shull, G.E.; Santoni, M.J.; Akeson, R.

J. Cell Biol. 105, 2335-2345, 1987

A/Title: Identification of a cDNA clone that contains the complete coding sequence for a

A/Reference number: S00846; MUID:8059265; PMID:3680385

A/Accession: S00846

A/Molecule type: mRNA

A/Residues: 1-858 &lt;SMA&gt;

A/Cross-references: EMBL:X06564

R/Small, S.J.; Akeson, R.

J. Cell Biol. 111, 2089-2096, 1990

A/Title: Expression of the unique NCAM VASE exon is independently regulated in distinct

A/Reference number: A37795; MUID:91035620; PMID:1699951

A/Accession: B37795

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: mRNA

A/Residues: 340-381 &lt;SM2&gt;

R/Small, S.J.; Haines, S.L.; Akeson, R.A.

Neuron 1, 1007-1017, 1988

A/Title: Polypeptide variation in an N-CAM extracellular immunoglobulin-like fold is dev

A/Reference number: I58136; MUID:90166485; PMID:2463093

A/Accession: I58136

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 355-364 &lt;RES&gt;

A/Cross-references: GB:M32611; NID:g205643; PIDN:AAA41679.1; PID:g205644

C/Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol

C/Genetics: Various forms of NCAM are produced by alternative splicing.

A/Status: NCM

C/Keywords: alternative splicing; brain; cell adhesion; duplication; heparin binding; s

F/1-19/Domain: signal sequence #status predicted &lt;SIG&gt;

F/20-721/Domain: extracellular #status predicted &lt;EXT&gt;

F/34-98/Domain: immunoglobulin homology &lt;IMM1&gt;

F/132-191/Domain: immunoglobulin homology &lt;IMM2&gt;

F/152-156/Region: heparin binding #status predicted

F/161-165/Region: heparin binding #status predicted

F/228-290/Domain: immunoglobulin homology &lt;IMM3&gt;

F/263-272/Region: NCAM binding #status predicted

F/323-398/Domain: immunoglobulin homology &lt;IMM4&gt;

F/430-492/Domain: immunoglobulin homology &lt;IMM5&gt;

F/529-606/Domain: fibronectin type III repeat homology &lt;FN3A&gt;

F/635-695/Domain: fibronectin type III repeat homology &lt;FN3B&gt;

F/722-739/Domain: transmembrane #status predicted &lt;TM&gt;

F/740-858/Domain: intracellular #status predicted &lt;INT&gt;

F/41-96/139-189,235-288,330-396,437-490/Disulfide bonds: #status predicted

F/222,316,348,434,460,489/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 52.2%; Score 47.5; DB 1; Length 858;

Best Local Similarity 60.0%; Pred. No. 11;

Matches 9; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

OY 3 PEOQETVLDGNLIR 17

Db 360 PEQET-LDGHMVR 373

## RESULT 7

hypothetical protein T6H20.280 - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 29-Oct-1999

C/Accession: T12981

R/Choisme, N.; Robert, C.; Brottier, P.; Wincker, P.; Catolico, L.; Artiguenave, F.; Sa

submitted to the Protein Sequence Database, July 1999

A/Reference number: Z17586

A/Accession: T12981

A/Molecule type: DNA

A/Residues: 1-452 &lt;CHO&gt;

A/Cross-references: EMBL:AL096859; GSPDB:GN00061; ATSP:T6H20.280

A/Experimental source: cultivar Columbia; BAC clone T6H20

C/Genetics:

A/Status: ATSP:T6H20.280

A/Map position: 3

A/Accession: T6H20.280

A/Title: Flavonol O3-glucosyltransferase

A/Reference number: 161/1

C/Supernatural: Flavonol O3-glucosyltransferase

Query Match 49.5%; Score 45; DB 2; Length 452;

Best Local Similarity 50.0%; Pred. No. 13;

Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 1 KSPEQETVLDGNLIR 18

Db 160 KDEKQDKVLEGLHPLRY 177

## RESULT 8

S-locus-specific glycoprotein precursor - rape

C/Species: Brassica napus (rape)

C/Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 20-Aug-1999

C/Accession: S25068; S19812

R/Goring, D.R.; Banks, P.; Beversdorf, W.D.; Rothstein, S.J.

Mol. Gen. Genet. 234, 185-192, 1992

A/Title: Use of the polymerase chain reaction to isolate an S-locus glycoprotein cDNA in

A/Reference number: S25068; MUID:92374986; PMID:1508146

A/Accession: S25068

A/Molecule type: mRNA

A/Residues: 1-409 &lt;GCR&gt;

A/Cross-references: EMBL:Z11724; NID:G17875; PIDN:CAAT7787.1; PID:G17876

C/Keywords: S-locus-specific glycoprotein; S-locus-specific glycoprotein homology

F/1-31/Domain: signal sequence #status predicted &lt;SIG&gt;

F/33-409/Domain: S-locus-specific glycoprotein #status predicted &lt;MAT&gt;

F/43,116,123,264,393/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 48.4%; Score 44; DB 2; Length 409;

Best Local Similarity 44.4%; Pred. No. 18;

Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 1 KSPEQETVLDGNLIR 18

Db 130 RSPVVALNGNFIIR 147

## RESULT 9

S-locus-specific glycoprotein - turnip (fragment)

N/Alternate names: S glycoprotein

C/Species: Brassica rapa (turnip)

C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 03-Nov-2000

C/Accession: T14415

R/Kisaba, M.; Nishio, T.; Satra, Y.; Hinata, K.; Ockendon, D.

Proc. Natl. Acad. Sci. U.S.A. 94, 7673-7678, 1997

A/Title: Striking sequence similarity in inter- and intra-specific comparisons of f class

A/Reference number: Z18078; MUID:97352858; PMID:9207151

A/Accession: T14415

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-431 &lt;KUS&gt;

A/Cross-references: EMBL:D95215; NID:g2351161; PIDN:BAA21949.1; PID:g2351162

C/Supernatural: S-locus-specific glycoprotein; S-locus-specific glycoprotein homology

C/Keywords: glycoprotein

Query Match 48.4%; Score 44; DB 2; Length 431;

Best Local Similarity 44.4%; Pred. No. 19;  
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 KSPQOETVLDGNLII 18  
: || | : || | : || |  
Db 122 RSPVVAEILANGNFVIRY 139

## RESULT 10

C49941  
devs protein - Myxococcus xanthus  
C/Species: Myxococcus xanthus  
C/Date: 19-May-1995 #sequence\_revision 19-May-1995 #text\_change 08-Oct-1999  
C/Accession: C49941  
R/Thoeny-Meyer, L.; Kaiser, D.  
J. Bacteriol. 175, 7450-7462, 1993  
A/Title: devRS, an autoregulated and essential genetic locus for fruiting body development  
A/Reference number: A49941; MUID:94042321; PMID:7693658  
A/Accession: C49941  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-214 <THO>  
A/Cross-references: GB:L19029; NID:g309547; PIDN:AAA16135.1; PID:g309549  
C/Genetics:  
A/Start codon: GTG

Query Match 47.3%; Score 43; DB 2; Length 214;  
Best Local Similarity 40.0%; Pred. No. 12;  
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 SPQOETVLDGNLII 16  
: || | : || | : || |  
Db 88 APDOOVLIDARLVV 102

## RESULT 11

T43683  
mRNA export factor mex67 - fission yeast (Schizosaccharomyces pombe)  
C/Species: Schizosaccharomyces pombe  
C/Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 18-Feb-2000  
C/Accession: T43683; T39787  
R/Yoon, J.H.; Whalen, W.A.; Bharathi, A.; Dhar, R.  
submitted to the EMBL Data Library, March 1998  
A/Description: Genetic and physical interaction between mRNA export factors Rexp and Me  
A/Reference number: 222630  
A/Accession: T43683  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-596 <YOO>  
A/Cross-references: EMBL:AF055036; PIDN:AAD43831.1  
A/Experimental source: Strain 972  
R/Seeger, K.; Harris, D.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, October 1999  
A/Reference number: 221816  
A/Accession: T39787  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-596 <SEB>  
A/Cross-references: EMBL:AL122033; PIDN:CAB58969.1; GSPDB:GN00067; SPDB:SPBC1921.03C  
C/Genetics:  
A/Experimental source: Strain 972h-, cosmid c1921  
A/Genes: mex67; SPBC1921.03C  
A/Map position: 2

Query Match 47.3%; Score 43; DB 2; Length 596;  
Best Local Similarity 62.5%; Pred. No. 41;  
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 KSPQOETVLDGNLII 16  
: || | : || | : || |  
Db 266 KLPELQELVAVGNPIV 281

## RESULT 12

1YH02  
inter-alpha-trypsin inhibitor heavy chain 2 precursor - human  
M/Alternate names: inter-alpha-trypsin inhibitor complex component II  
C/Species: Homo sapiens (man)  
C/Date: 30-Sep-1990 #sequence\_revision 30-Sep-1990 #text\_change 04-Feb-2000  
C/Accession: S00346; S09064; E39967; S00632; B34245; S28929; C53642  
R/Gebhard, W.; Schreitmuller, T.; Hochstrasser, K.; Wachter, E.  
FEBS Lett. 229, 63-67, 1988  
A/Title: Complementary DNA and derived amino acid sequence of the precursor of one of the  
A/Reference number: S00346; MUID:88152237; PMID:2450046  
A/Accession: S00346  
A/Molecule type: mRNA  
A/Residues: 1-946 <GEB>  
A/Cross-references: EMBL:X07173  
A/Experimental source: liver  
A/Note: part of this sequence, including the amino end of the mature protein, was confirmed  
A/Note: due to a double frameshift, the nucleic acid sequence of codons 363-372 is incorrect  
A/Note: in one clone, a T is lacking from codon 716; this clone could code for a protein  
R/Schreitmuller, T.; Hochstrasser, K.; Reisinger, P.W.M.; Wachter, E.; Gebhard, W.  
Biol. Chem. Hoppe-Seyler 368, 963-970, 1987  
A/Title: cDNA cloning of human inter-alpha-trypsin inhibitor discloses three different pr  
A/Reference number: S09064; MUID:88024442; PMID:3663330  
A/Accession: S09064  
A/Molecule type: mRNA  
A/Residues: 265; 'RR', 268-284, 'D', 286-946 <SCH>  
A/Note: this sequence has been revised in reference S00346  
R/Salter, J.P.; Diarra-Mehrpour, M.; Sesboue, R.; Bouguignon, J.; Benarous, R.; Ohkubo,  
Proc. Natl. Acad. Sci. U.S.A. 84, 8272-8276, 1987  
A/Title: Isolation and characterization of cDNAs encoding the heavy chain of human inter-  
A/Reference number: A39967; MUID:88068576; PMID:2446322  
A/Accession: B39967  
A/Molecule type: mRNA  
A/Residues: 384-673 'A', 675-704 'S', 706-728 'D', 730 'A', 732-865 <SAL>  
A/Cross-references: GB:M8183; GB:030313; NID:g338222; PIDN:AAA60558.1; PID:g553647  
A/Experimental source: liver  
R/Salter, J.P.; Diarra-Mehrpour, M.; Sesboue, R.; Bouguignon, J.; Martin, J.P.  
Biol. Chem. Hoppe-Seyler 369 (Suppl.), 15-18, 1988  
A/Title: Human inter-alpha-trypsin inhibitor. Isolation and characterization of heavy (H)  
A/Reference number: S00632; MUID:89076497; PMID:2462430  
A/Accession: S00632  
A/Molecule type: mRNA  
A/Residues: 384-673 'A', 675-704 'S', 706-728 'D', 730 'A', 732-766 <SAS>  
A/Cross-references: GB:M33033; NID:g186589; PIDN:AAA59195.1; PID:g186590  
R/Englund, J.U.; Thøgersen, I.B.; Pizzo, S.V.; Salvesen, G.  
J. Biol. Chem. 264, 15975-15981, 1989  
A/Title: Analysis of inter-alpha-trypsin inhibitor and a novel trypsin inhibitor, pre-  
A/Reference number: A92736; MUID:89380192; PMID:2476436  
A/Accession: B34245  
A/Molecule type: protein  
A/Residues: 55-74 <ENG>  
R/Maki, N.; Baldyck, M.; Maes, P.; Capon, C.; Mison, C.; Han, K.K.; Tartar, A.; Fournet  
Biol. Chem. Hoppe-Seyler 373, 1009-1018, 1992  
A/Title: The heavy chains of human plasma inter-alpha-trypsin inhibitor: their isolation,  
A/Reference number: S28928; MUID:93039735; PMID:1384548  
A/Accession: S28929  
A/Molecule type: protein  
A/Residues: 55-64 <MAL>  
R/Wisniewski, H.G.; Burgess, W.H.; Oppenheim, J.D.; Vilecek, J.  
Biochemistry 33, 7423-7429, 1994  
A/Title: TSG-6, an arthritis-associated hyaluronan binding protein, forms a stable complex  
A/Reference number: A53642; MUID:94271799; PMID:7516184  
A/Accession: C53642  
A/Molecule type: protein  
A/Residues: 55-64 <WIS>  
C/Comment: Inter-alpha-trypsin inhibitor is a complex of three proteins, each deriving f  
C/Genetics: This protein is a heterodimer of heavy and light chains.  
A/Genes: GDB:ITIH2  
A/Cross-references: GDB:120108; OMIM:146640  
A/Map position: 10p15-10p15  
C/Superfamily: inter-alpha-trypsin inhibitor complex component II

C;Keywords: carboxylglutamic acid; glycoprotein; heterodimer; serine proteinase inhibitor  
 F;1-18/Domain: signal sequence #status predicted <SIG>  
 F;19-54/Domain: propeptide #status predicted <PRO>  
 F;55-698/Protein: inter-alpha-trypsin inhibitor heavy chain 2 #status predicted <MAT>  
 F;96-445/Binding site: carbohydrate (asn) (covalent) #status predicted  
 F;118,671/Binding site: carbohydrate (asn) (covalent) #status predicted  
 F;282,283/Modified site: gamma-carboxylglutamic acid (gu) #status predicted  
 F;421,422,423/Binding site: calcium (asp, gly, asp) #status predicted

Query Match 47.3%; Score 43; DB 1; Length 946;  
 Best Local Similarity 43.8%; Pred. No. 70;  
 Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 3 PEOQETVLDGNLIIRY 18  
 Db 262 PSCHETAVDGLVLY 277

RESULT 13  
 T20389  
 hypochelical protein DY3.1 - Caenorhabditis elegans  
 C;Species: Caenorhabditis elegans  
 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C;Accession: T20389  
 R;Lennard, N.  
 submitted to the EMBL Data Library, June 1997  
 A;Reference number: Z19266  
 A;Accession: T20389  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-108 <WIL>  
 A;Cross-references: EMBL:Z96047; PIDN:CA809410.1; GSPDB:GN00019; CESP:DY3.1  
 C;Experimental source: clone DY3  
 C;Genetics:  
 A;Gene: CESP:DY3.1  
 A;Map position: 1  
 A;Introns: 39/3

Query Match 46.2%; Score 42; DB 2; Length 108;  
 Best Local Similarity 72.7%; Pred. No. 8.1;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SPEQETVLDG 12  
 Db 14 SPEQEQVIG 24

RESULT 14  
 D69588  
 transcription repressor of arabinose operon arar - Bacillus subtilis  
 C;Species: Bacillus subtilis  
 C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
 C;Accession: D69588  
 R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertet  
 C.; Bron, S.; Brouillet, S.; Busschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch  
 A.; Ehrlich, S.D.; Emmerston, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gall  
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapides, A.; Lardinois  
 A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Potte  
 Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,  
 A.; Authors: Schleich, S.; Schroeter, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
 akehchi, M.; Yamakoshi, A.; Tanaka, T.; Terstegen, P.; Togonou, K.; Tosato, V.; Uchiyama,  
 T.; Witter, P.; Wipet, A.; Yamamoto, H.; Yamano, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
 A.; Authors: Yoshikawa, H.F.; Zundstein, E.; Yoshikawa, H.; Danchin, A.  
 A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A;Reference number: A69580; MUID:98044033; PMID:9384377  
 A;Accession: D69588  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-384 <KUN>

A;Cross-references: GB:Z99121; GB:AL009126; NID:G2635827; PIDN:CAB15402.1; PID:e1186085;  
 A;Experimental source: strain 168  
 C;Genetics:  
 A;Gene: arar

Query Match 46.2%; Score 42; DB 2; Length 384;  
 Best Local Similarity 41.2%; Pred. No. 36;  
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 1 KSPQETVLDGNLIIR 17  
 Db 360 KKPQEDVIEPELIIR 376

RESULT 15  
 T14419  
 S-locus-specific glycoprotein - turnip (fragment)  
 N;Alternate names: S glycoprotein  
 C;Species: Brassica rapa (turnip)  
 C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 03-Nov-2000  
 C;Accession: T14419  
 R;Kusab, M.; Nishio, T.; Satta, Y.; Hinata, K.; Ockendon, D.  
 Proc. Natl. Acad. Sci. U.S.A. 94, 7673-7678, 1997  
 A;Title: Striking sequence similarity in inter- and intra-specific comparisons o f class  
 echanism.  
 A;Reference number: Z18078; MUID:97352858; PMID:9207151  
 A;Accession: T14419  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-428 <KUS>  
 A;Cross-references: EMBL:D85219; NID:G2351169; PIDN:BAA21953.1; PID:G2351170  
 C;Superfamily: S-locus-specific glycoprotein; S-locus-specific glycoprotein homology  
 C;Keywords: glycoprotein

Query Match 46.2%; Score 42; DB 2; Length 428;  
 Best Local Similarity 44.4%; Pred. No. 41;  
 Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 1 KSPQETVLDGNLIIR 18  
 Db 120 RSPVALLPQGNFVIRY 137

RESULT 16  
 S19023  
 chaperonin groEL - Chlamydia pneumoniae  
 N;Alternate names: heat shock protein-60  
 C;Species: Chlamydia pneumoniae  
 C;Date: 22-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 31-Mar-2000  
 C;Accession: S19023; C72114  
 R;Kikuta, L.C.; Puolakkainen, M.; Kuo, C.C.; Campbell, L.A.  
 Infect. Immun. 59, 4665-4669, 1991  
 A;Title: Isolation and sequence analysis of the Chlamydia pneumoniae GroE operon.  
 A;Reference number: S19022; MUID:92040169; PMID:1682266  
 A;Accession: S19023  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-544 <KIK>  
 A;Cross-references: EMBL:X60068; NID:G48931; PIDN:CAA42673.1; PID:G48933  
 R;Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.  
 Nature Genet. 21, 385-389, 1999  
 A;Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
 A;Reference number: A72000; MUID:99206606; PMID:10192388  
 A;Accession: C72114  
 A;Molecule type: DNA  
 A;Residues: 1-509, 'A', 511-544 <ARN>  
 A;Cross-references: GB:AE001600; GB:AE001363; NID:G4376398; PIDN:AAD18287.1; PID:G4376395;  
 A;Experimental source: strain CW029  
 C;Genetics:  
 A;Gene: groEL\_1  
 C;Superfamily: chaperonin groEL

Query Match 46.2%; Score 42; DB 2; Length 544;



A:Residues: 1003-1285 <ROY>  
A:Cross-references: EMBL:X59371  
C:Genetics:  
A:Gene: MIP5:YKL215C  
A:Cross-references: SGD:S0001698  
A:Map position: 11L  
C:Superfamily: hypothetical protein YKL215C

Query Match 46.2% Score 42; DB 1; Length 1286;  
Best Local Similarity 58.3%; Pred. No. 1.5e+02;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 2 SP00EYLDGN 13  
Db 156 SP00EGILEGN 167

RESULT 21  
BVECAI  
DNA helicase I (EC 3.6.1.-) - Escherichia coli plasmids  
C:Species: Escherichia coli  
C:Date: 30-Jun-1990 #sequence revision 21-Jul-2000 #text change 19-Jan-2001  
C:Accession: S10660; PS0322; PS0068; T00299; T42198  
R:Yoshida, Y.; Fujita, Y.; Ohtsubo, E.  
J. Mol. Biol. 214, 39-53, 1990  
A:Title: Nucleotide sequence of the promoter-distal region of the tra operon of plasmid  
A:Reference number: S10658; MUID:90317835; PMID:2164585  
A:Accession: S10660  
A:Molecule type: DNA  
A:Residues: 1-1756 <YOS>  
A:Cross-references: EMBL:X5815; NID:g42620; PIDN:CAA3937.1; PID:g42623  
A:Experimental source: plasmid R100  
R:Cram, D.S.; Ioh, S.M.; Cheah, K.C.; Skurray, R.A.  
Gene 104, 85-90, 1991  
A:Title: Sequence and conservation of genes at the distal end of the transfer region on  
A:Reference number: J01338; MUID:92009201; PMID:13916281  
A:Accession: PS0322  
A:Molecule type: DNA  
A:Residues: 1747-1756 <CRA1>  
A:Cross-references: GB:M38048  
A:Experimental source: plasmid R6-5  
A:Accession: PS0321  
A:Molecule type: DNA  
A:Residues: 1747-1756 <CRA2>  
A:Cross-references: GB:M38047; NID:g148651; PIDN:AAA98090.1; PID:g148652  
A:Experimental source: plasmid F  
R:Jalajakumari, M.B.; Manning, P.A.  
Gene 81, 195-202, 1989  
A:Title: Nucleotide sequence of the traD region in the Escherichia coli F sex factor.  
A:Reference number: J50293; MUID:90034191; PMID:2680768  
A:Accession: PS0068  
A:Molecule type: DNA  
A:Residues: 1,'M','3','IAQ','7','R','9-35','R','37-68','CRNAVTR','76-150 <JAL>  
A:Cross-references: GB:M29254; NID:g148618; PIDN:AAA83930.1; PID:g551859  
A:Experimental source: strain K12; F factor  
A:Note: the authors translated the codon CTG for residue 41 as Glu  
R:Makino, K.; Ishii, K.; Yasunaga, T.; Hattori, M.; Yokoyama, K.; Yatsudo, H.C.; Kubota,  
S.; Shinagawa, H.  
DNA Res. 5, 1-9, 1998  
A:Title: Complete nucleotide sequences of 93-kb and 3.3-kb plasmids of an enterohemorrhagic  
A:Reference number: Z14127; MUID:98290540; PMID:9628576  
A:Accession: T00299  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 'S','1108-1109','E','1111','G','1113','FN','1295-1301','R','1303-1309','Q','1311-1332','  
'N','1632-1682','R','1684-1696','E','1698-1699','VTS','1703-1715','P','1717-1719','R','1721-1726','V',  
A:Cross-references: EMBL:AB011549; PIDN:BA31818.1  
A:Experimental source: strain EHEC O157:H7, substrain RIMD 0509952  
R:Burland, V.; Shao, Y.; Perna, N.T.; Plunkett, G.; Sofia, H.J.; Blattner, F.R.  
Nucleic Acids Res. 26, 4196-4204, 1998  
A:Title: The complete DNA sequence and analysis of the large virulence plasmid of Escherichia  
A:Reference number: Z22068; MUID:98391744; PMID:9722640  
A:Accession: T42198

```

A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 'M',126-1332,'P',1334-1335,'T',1337-1352,'G',1354-1375,'G',1377-1385,'A',1389-1390,'C',1391-1400,'U',1401-1402,'V',1403-1404,'I',1405-1406,'L',1407-1408,'S',1409-1410,'K',1411-1412,'R',1413-1414,'V',1415-1416,'W',1417-1418,'M',1419-1420,'I',1421-1422,'V',1423-1424,'V',1425-1426,'V',1427-1428,'V',1429-1430,'V',1431-1432,'V',1433-1434,'V',1435-1436,'V',1437-1438,'V',1439-1440,'V',1441-1442,'V',1443-1444,'V',1445-1446,'V',1447-1448,'V',1449-1450,'V',1451-1452,'V',1453-1454,'V',1455-1456,'V',1457-1458,'V',1459-1460,'V',1461-1462,'V',1463-1464,'V',1465-1466,'V',1467-1468,'V',1469-1470,'V',1471-1472,'V',1473-1474,'V',1475-1476,'V',1477-1478,'V',1479-1480,'V',1481-1482,'V',1483-1484,'V',1485-1486,'V',1487-1488,'V',1489-1490,'V',1491-1492,'V',1493-1494,'V',1495-1496,'V',1497-1498,'V',1499-1500,'V',1501-1502,'V',1503-1504,'V',1505-1506,'V',1507-1508,'V',1509-1510,'V',1511-1512,'V',1513-1514,'V',1515-1516,'V',1517-1518,'V',1519-1520,'V',1521-1522,'V',1523-1524,'V',1525-1526,'V',1527-1528,'V',1529-1530,'V',1531-1532,'V',1533-1534,'V',1535-1536,'V',1537-1538,'V',1539-1540,'V',1541-1542,'V',1543-1544,'V',1545-1546,'V',1547-1548,'V',1549-1550,'V',1551-1552,'V',1553-1554,'V',1555-1556,'V',1557-1558,'V',1559-1560,'V',1561-1562,'V',1563-1564,'V',1565-1566,'V',1567-1568,'V',1569-1570,'V',1571-1572,'V',1573-1574,'V',1575-1576,'V',1577-1578,'V',1579-1580,'V',1581-1582,'V',1583-1584,'V',1585-1586,'V',1587-1588,'V',1589-1590,'V',1591-1592,'V',1593-1594,'V',1595-1596,'V',1597-1598,'V',1599-1600,'V',1601-1602,'V',1603-1604,'V',1605-1606,'V',1607-1608,'V',1609-1610,'V',1611-1612,'V',1613-1614,'V',1615-1616,'V',1617-1618,'V',1619-1620,'V',1621-1622,'V',1623-1624,'V',1625-1626,'V',1627-1628,'V',1629-1630,'V',1631-1632,'V',1633-1634,'V',1635-1636,'V',1637-1638,'V',1639-1640,'V',1641-1642,'V',1643-1644,'V',1645-1646,'V',1647-1648,'V',1649-1650,'V',1651-1652,'V',1653-1654,'V',1655-1656,'V',1657-1658,'V',1659-1660,'V',1661-1662,'V',1663-1664,'V',1665-1666,'V',1667-1668,'V',1669-1670,'V',1671-1672,'V',1673-1674,'V',1675-1676,'V',1677-1678,'V',1679-1680,'V',1681-1682,'V',1683-1684,'V',1685-1686,'V',1687-1688,'V',1689-1690,'V',1691-1692,'V',1693-1694,'V',1695-1696,'V',1697-1698,'V',1699-1700,'V',1701-1702,'V',1703-1704,'V',1705-1706,'V',1707-1708,'V',1709-1710,'V',1711-1712,'V',1713-1714,'V',1715-1716,'V',1717-1718,'V',1719-1720,'V',1721-1722,'V',1723-1724,'V',1725-1726,'V',1727-1728,'V',1729-1730,'V',1731-1732,'V',1733-1734,'V',1735-1736,'V',1737-1738,'V',1739-1740,'V',1741-1742,'V',1743-1744,'V',1745-1746,'V',1747-1748,'V',1749-1750,'V',1751-1752,'V',1753-1754,'V',1755-1756,'V',1757-1758,'V',1759-1760,'V',1761-1762,'V',1763-1764,'V',1765-1766,'V',1767-1768,'V',1769-1770,'V',1771-1772,'V',1773-1774,'V',1775-1776,'V',1777-1778,'V',1779-1780,'V',1781-1782,'V',1783-1784,'V',1785-1786,'V',1787-1788,'V',1789-1790,'V',1791-1792,'V',1793-1794,'V',1795-1796,'V',1797-1798,'V',1799-1800,'V',1801-1802,'V',1803-1804,'V',1805-1806,'V',1807-1808,'V',1809-1810,'V',1811-1812,'V',1813-1814,'V',1815-1816,'V',1817-1818,'V',1819-1820,'V',1821-1822,'V',1823-1824,'V',1825-1826,'V',1827-1828,'V',1829-1830,'V',1831-1832,'V',1833-1834,'V',1835-1836,'V',1837-1838,'V',1839-1840,'V',1841-1842,'V',1843-1844,'V',1845-1846,'V',1847-1848,'V',1849-1850,'V',1851-1852,'V',1853-1854,'V',1855-1856,'V',1857-1858,'V',1859-1860,'V',1861-1862,'V',1863-1864,'V',1865-1866,'V',1867-1868,'V',1869-1870,'V',1871-1872,'V',1873-1874,'V',1875-1876,'V',1877-1878,'V',1879-1880,'V',1881-1882,'V',1883-1884,'V',1885-1886,'V',1887-1888,'V',1889-1890,'V',1891-1892,'V',1893-1894,'V',1895-1896,'V',1897-1898,'V',1899-1900,'V',1901-1902,'V',1903-1904,'V',1905-1906,'V',1907-1908,'V',1909-1910,'V',1911-1912,'V',1913-1914,'V',1915-1916,'V',1917-1918,'V',1919-1920,'V',1921-1922,'V',1923-1924,'V',1925-1926,'V',1927-1928,'V',1929-1930,'V',1931-1932,'V',1933-1934,'V',1935-1936,'V',1937-1938,'V',1939-1940,'V',1941-1942,'V',1943-1944,'V',1945-1946,'V',1947-1948,'V',1949-1950,'V',1951-1952,'V',1953-1954,'V',1955-1956,'V',1957-1958,'V',1959-1960,'V',1961-1962,'V',1963-1964,'V',1965-1966,'V',1967-1968,'V',1969-1970,'V',1971-1972,'V',1973-1974,'V',1975-1976,'V',1977-1978,'V',1979-1980,'V',1981-1982,'V',1983-1984,'V',1985-1986,'V',1987-1988,'V',1989-1990,'V',1991-1992,'V',1993-1994,'V',1995-1996,'V',1997-1998,'V',1999-2000,'V',2001-2002,'V',2003-2004,'V',2005-2006,'V',2007-2008,'V',2009-2010,'V',2011-2012,'V',2013-2014,'V',2015-2016,'V',2017-2018,'V',2019-2020,'V',2021-2022,'V',2023-2024,'V',2025-2026,'V',2027-2028,'V',2029-2030,'V',2031-2032,'V',2033-2034,'V',2035-2036,'V',2037-2038,'V',2039-2040,'V',2041-2042,'V',2043-2044,'V',2045-2046,'V',2047-2048,'V',2049-2050,'V',2051-2052,'V',2053-2054,'V',2055-2056,'V',2057-2058,'V',2059-2060,'V',2061-2062,'V',2063-206
```

## RESULT 26

RESULT 28  
JC2250

C:\Accession: JC2250; PC2166; D26424  
R.Yamakawa, S., Shiba, H.; Watanabe, M.; Shiozawa, H.; Takayama, S.; Hinata, K.; Isegai, Biosci. Biotechnol. Biochem. 58, 921-925, 1994

A;Title: The sequences of S-glycoproteins involved in self-incompatibility of Brassica  
A;Reference number: JC2250; MUID:9428987; PMID:7764979  
A;Accession: JC2250  
A;Molecule type: mRNA  
A;Residues: 1-436 <YAM>  
A;Accession: PC2166  
A;Molecule type: protein  
A;Residues: 32-60;78-100;102-152;172-198;202-261;305-346;364-399 <YA2>  
R;Takayama, S.; Isogai, A.; Tsukamoto, C.; Ueda, Y.; Hinata, K.; Okazaki, K.; Suzuki, A.  
Nature 326, 102-105, 1987  
A;Title: Sequences of S-glycoproteins, products of the Brassica campestris self-incompat  
A;Reference number: A93390  
A;Accession: D26424  
A;Molecule type: protein  
A;Residues: 32-45; 'X', 47-60;78-100;102-110; 'X', 112-113; 'X', 115;114-120; 'XI', 123-152;172-  
C;Comment: This protein plays a role in self-incompatibility in Brassica plants. The sel  
C;Superfamily: S-locus-specific glycoprotein; S-locus-specific glycoprotein homology  
C;Keywords: glycoprotein  
F;1-31/Domain: signal sequence #status predicted <SIG>  
F;32-45/Product: S-locus-specific glycoprotein S12 #status experimental <MAT>  
F;40-434/Domain: S-locus-specific glycoprotein homology <SSG>  
F;46,121,246/Binding site: carbohydrate (Asn) (covalent) #status experimental  
Query Match 45.1%; Score 41; DB 1; Length 436;  
Best Local Similarity 47.1%; Pred. No. 62;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
QY 1 KSPQQTVDGNLIR 17  
DB 128 RSPVLAELANGNLVIR 144  
RESULT 29  
T24478  
hypothetical protein T04H1.8 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Mar-2000  
C;Accession: T24478  
R;Harris, B.  
submitted to the EMBL Data Library, August 1996  
A;Reference number: Z19896  
A;Accession: T24478  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-440 <WTL>  
A;Cross-references: EMBL:Z78200; PIDD: CAB01585.1; GSPDB: GNO0023; CESP: T04H1.8  
A;Experimental source: clone T04H1  
C;Gene: CESP: T04H1.8  
A;Map position: 5  
A;Intons: 50/3; 106/3; 187/1; 280/3; 359/1  
C;Superfamily: glucuronosyltransferase  
Query Match 45.1%; Score 41; DB 2; Length 440;  
Best Local Similarity 54.5%; Pred. No. 63;  
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
QY 2 SPEGQETVLDG 12  
DB 100 SPEGQRTMEG 110  
RESULT 30  
A64117  
serine-type D-Ala-D-Ala carboxypeptidase homolog - Haemophilus influenzae (strain Rd KW2  
C;Species: Haemophilus influenzae  
C;Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 17-Mar-2000  
C;Accession: A64117  
R;Reichmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,

A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A;Reference number: A64000; MUID:9530630; PMID:7542800  
A;Accession: A64117  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-479 <TIGR>  
A;Cross-references: GB:U32812; GB:I42023; NID:g1574784; PIDD: AAC22975.1; PID:g1574789; TJ  
C;Superfamily: D-alanyl-D-alanine carboxypeptidase  
Query Match 45.1%; Score 41; DB 2; Length 479;  
Best Local Similarity 77.8%; Pred. No. 69;  
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 10 LDGNLIRY 18  
DB 104 LDGNLIRY 112  
RESULT 31  
A71555  
probable hsp-60 - Chlamydia trachomatis (serotype D, strain UM3/Cx)  
C;Species: Chlamydia trachomatis  
C;Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 26-Aug-1999  
C;Accession: A71555  
R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,  
Science 282, 754-759, 1998  
A;Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tract  
A;Reference number: A71570; MUID:9900809; PMID:9784136  
A;Accession: A71555  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-544 <ARN>  
A;Cross-references: GB:AE001285; GB:AE001273; NID:g3328507; PIDD: AAC67701.1; PID:g3328508  
A;Experimental source: serotype D, strain UM-3/Cx  
C;Gene: groEL  
C;Superfamily: chaperonin groEL  
Query Match 45.1%; Score 41; DB 2; Length 544;  
Best Local Similarity 47.1%; Pred. No. 80;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
QY 2 SPEGQETVLDGNLIRY 18  
DB 207 NPETQECVLEDAVLTY 223  
RESULT 32  
I40731  
heat shock protein - Chlamydia trachomatis  
C;Species: Chlamydia trachomatis  
C;Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 26-Aug-1999  
C;Accession: I40731  
R;Ho, Y.; Zhang, Y.X.  
Gene 141, 143-144, 1994  
A;Title: The sequence of the groES and groEL genes from the mouse pneumonitis agent of Ch  
A;Reference number: I40730; MUID:94215880; PMID:7909303  
A;Accession: I40731  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-544 <RES>  
A;Cross-references: GB:I12004; NID:g402332; PIDD: AAA19871.1; PID:g402333  
C;Gene: groEL  
C;Superfamily: chaperonin groEL  
Query Match 45.1%; Score 41; DB 2; Length 544;  
Best Local Similarity 47.1%; Pred. No. 80;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
QY 2 SPEGQETVLDGNLIRY 18  
DB 207 NPETQECVLEDAVLTY 223



RESULT 33

B41479

60K heat shock protein protein groEL - Chlamydia trachomatis

N|Alternative names: stress response protein HypB

C|Species: Chlamydia trachomatis

C|Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Oct-1999

C|Accession: B41479; A60273

R|Morrison, R.P.; Su, H.; Lyng, K.; Yuan, Y.

Infect. Immun. 58, 2701-2705, 1990

A|Title: The Chlamydia trachomatis hyp operon is homologous to the groE stress response

A|Reference number: A41479; MUID:90316704; PMID:2196231

A|Accession: B41479

A|Molecule type: DNA

A|Residues: 1-544 <MOR>

A|Cross-references: GB:M31739; NID:G144520; PID:AAA03204.1; PID:G144522

R|Cerrone, M.C.; Ma, J.U.; Stephens, R.S.

Infect. Immun. 59, 79-90, 1991

A|Title: Cloning and sequence of the gene for heat shock protein 60 from Chlamydia trachomatis

A|Reference number: A60273; MUID:9110013; PMID:1987066

A|Accession: A60273

A|Molecule type: DNA

A|Residues: 1-123, 'A', 125-130, 'V', 132-190, 'D', 192-216, 'E', 218-235, 'V', 237-254, 'V', 256-282, 'V', 284-300, 'V', 302-312, 'V', 314-324, 'V', 326-336, 'V', 338-348, 'V', 350-360, 'V', 362-372, 'V', 374-384, 'V', 386-396, 'V', 398-408, 'V', 410-420, 'V', 422-432, 'V', 434-444, 'V', 446-456, 'V', 458-468, 'V', 470-480, 'V', 482-492, 'V', 494-504, 'V', 506-516, 'V', 518-528, 'V', 530-540, 'V', 542-552, 'V', 554-564, 'V', 566-576, 'V', 578-588, 'V', 590-600, 'V', 602-612, 'V', 614-624, 'V', 626-636, 'V', 638-648, 'V', 650-660, 'V', 662-672, 'V', 674-684, 'V', 686-696, 'V', 698-708, 'V', 710-720, 'V', 722-732, 'V', 734-744, 'V', 746-756, 'V', 758-768, 'V', 770-780, 'V', 782-792, 'V', 794-804, 'V', 806-816, 'V', 818-828, 'V', 830-840, 'V', 842-852, 'V', 854-864, 'V', 866-876, 'V', 878-888, 'V', 890-900, 'V', 902-912, 'V', 914-924, 'V', 926-936, 'V', 938-948, 'V', 950-960, 'V', 962-972, 'V', 974-984, 'V', 986-996, 'V', 998-1008, 'V', 1010-1020, 'V', 1022-1032, 'V', 1034-1044, 'V', 1046-1056, 'V', 1058-1068, 'V', 1070-1080, 'V', 1082-1092, 'V', 1094-1104, 'V', 1106-1116, 'V', 1118-1128, 'V', 1130-1140, 'V', 1142-1152, 'V', 1154-1164, 'V', 1166-1176, 'V', 1178-1188, 'V', 1190-1200, 'V', 1202-1212, 'V', 1214-1224, 'V', 1226-1236, 'V', 1238-1248, 'V', 1250-1260, 'V', 1262-1272, 'V', 1274-1284, 'V', 1286-1296, 'V', 1298-1308, 'V', 1310-1320, 'V', 1322-1332, 'V', 1334-1344, 'V', 1346-1356, 'V', 1358-1368, 'V', 1370-1380, 'V', 1382-1392, 'V', 1394-1404, 'V', 1406-1416, 'V', 1418-1428, 'V', 1430-1440, 'V', 1442-1452, 'V', 1454-1464, 'V', 1466-1476, 'V', 1478-1488, 'V', 1490-1500, 'V', 1502-1512, 'V', 1514-1524, 'V', 1526-1536, 'V', 1538-1548, 'V', 1550-1560, 'V', 1562-1572, 'V', 1574-1584, 'V', 1586-1596, 'V', 1598-1608, 'V', 1610-1620, 'V', 1622-1632, 'V', 1634-1644, 'V', 1646-1656, 'V', 1658-1668, 'V', 1670-1680, 'V', 1682-1692, 'V', 1694-1704, 'V', 1706-1716, 'V', 1718-1728, 'V', 1730-1740, 'V', 1742-1752, 'V', 1754-1764, 'V', 1766-1776, 'V', 1778-1788, 'V', 1790-1800, 'V', 1802-1812, 'V', 1814-1824, 'V', 1826-1836, 'V', 1838-1848, 'V', 1850-1860, 'V', 1862-1872, 'V', 1874-1884, 'V', 1886-1896, 'V', 1898-1908, 'V', 1910-1920, 'V', 1922-1932, 'V', 1934-1944, 'V', 1946-1956, 'V', 1958-1968, 'V', 1970-1980, 'V', 1982-1992, 'V', 1994-2004, 'V', 2006-2016, 'V', 2018-2028, 'V', 2030-2040, 'V', 2042-2052, 'V', 2054-2064, 'V', 2066-2076, 'V', 2078-2088, 'V', 2090-2100, 'V', 2102-2112, 'V', 2114-2124, 'V', 2126-2136, 'V', 2138-2148, 'V', 2150-2160, 'V', 2162-2172, 'V', 2174-2184, 'V', 2186-2196, 'V', 2198-2208, 'V', 2210-2220, 'V', 2222-2232, 'V', 2234-2244, 'V', 2246-2256, 'V', 2258-2268, 'V', 2270-2280, 'V', 2282-2292, 'V', 2294-2304, 'V', 2306-2316, 'V', 2318-2328, 'V', 2330-2340, 'V', 2342-2352, 'V', 2354-2364, 'V', 2366-2376, 'V', 2378-2388, 'V', 2390-2400, 'V', 2402-2412, 'V', 2414-2424, 'V', 2426-2436, 'V', 2438-2448, 'V', 2450-2460, 'V', 2462-2472, 'V', 2474-2484, 'V', 2486-2496, 'V', 2498-2508, 'V', 2510-2520, 'V', 2522-2532, 'V', 2534-2544, 'V', 2546-2556, 'V', 2558-2568, 'V', 2570-2580, 'V', 2582-2592, 'V', 2594-2604, 'V', 2606-2616, 'V', 2618-2628, 'V', 2630-2640, 'V', 2642-2652, 'V', 2654-2664, 'V', 2666-2676, 'V', 2678-2688, 'V', 2690-2700, 'V', 2702-2712, 'V', 2714-2724, 'V', 2726-2736, 'V', 2738-2748, 'V', 2750-2760, 'V', 2762-2772, 'V', 2774-2784, 'V', 2786-2796, 'V', 2798-2808, 'V', 2810-2820, 'V', 2822-2832, 'V', 2834-2844, 'V', 2846-2856, 'V', 2858-2868, 'V', 2870-2880, 'V', 2882-2892, 'V', 2894-2904, 'V', 2906-2916, 'V', 2918-2928, 'V', 2930-2940, 'V', 2942-2952, 'V', 2954-2964, 'V', 2966-2976, 'V', 2978-2988, 'V', 2990-3000, 'V', 3002-3012, 'V', 3014-3024, 'V', 3026-3036, 'V', 3038-3048, 'V', 3050-3060, 'V', 3062-3072, 'V', 3074-3084, 'V', 3086-3096, 'V', 3098-3108, 'V', 3110-3120, 'V', 3122-3132, 'V', 3134-3144, 'V', 3146-3156, 'V', 3158-3168, 'V', 3170-3180, 'V', 3182-3192, 'V', 3194-3204, 'V', 3206-3216, 'V', 3218-3228, 'V', 3230-3240, 'V', 3242-3252, 'V', 3254-3264, 'V', 3266-3276, 'V', 3278-3288, 'V', 3290-3300, 'V', 3302-3312, 'V', 3314-3324, 'V', 3326-3336, 'V', 3338-3348, 'V', 3350-3360, 'V', 3362-3372, 'V', 3374-3384, 'V', 3386-3396, 'V', 3398-3408, 'V',

```

C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #ext_change 01-Mar-2002
C:Accession: E64820
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: E64820
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-612 <BLAT>
A:Cross-references: GB:AB000185; GB:U00096; NID:g1787047; PIDN:AACT3916.1; PID:g1787051;
C:Supplemental source: strain K-12, substrain MG1655
C:Keywords: ATP, nucleotide binding, P-loop
F:21-234/Domain: ATP-binding cassette homology <ABC1>
F:38-45/Region: nucleotide-binding motif A (P-loop)
F:329-529/Domain: ATP-binding cassette homology <ABC2>
F:346-353/Region: nucleotide-binding motif A (P-loop)

Query Match 45.1% Score 41; DB 2; Length 612;
Best Local Similarity 46.7% Pred. No. 92;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 3 PEQGTVDGNLIIR 17
|:|:|:|:|:|
DB 290 PIEKTVVDGEPVLR 304

RESULT 36
D00742
Hypothetical protein Ecs0908 [imported] - Escherichia coli (strain O157:H7, substrain RIT
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #ext_change 17-May-2002
C:Accession: D90742
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.,
Gasanawa, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
Nucleic Acids Res. 29, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
A:Reference number: A39629; MUID:21156231; PMID:11258796
A:Accession: D90742
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-612 <HAY>
A:Cross-references: PIDN:BA000007; PIDN:BA034331.1; PID:g13360367; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
A:Genetics:
A:Gene: Ecs0908
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 45.1% Score 41; DB 2; Length 612;
Best Local Similarity 46.7% Pred. No. 92;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 3 PEQGTVDGNLIIR 17
|:|:|:|:|:|
DB 290 PIEKTVVDGEPVLR 304

RESULT 37
G85592
Hypothetical protein Z1053 [imported] - Escherichia coli (strain O157:H7, substrain EDLg:
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #ext_change 17-May-2002
C:Accession: G85592
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobbeck, B.J.; Davis, N.W.; Linn, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G85592
A>Status: preliminary

```





**This Page Blank (uspio)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 10, 2003, 17:12:24 ; Search time 23 Seconds

(without alignments)  
36.804 Million cell updates/sec

Title: US-09-991-795-1

Perfect score: 91

Sequence: 1 KSPDQOETVLDGMLIRY 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91	100.0	930	1 ITH4_HUMAN	Q14624 h inter- $\alpha$ -
2	86	94.5	921	1 ITH4_PIG	P79263 sus scrofa
3	47.5	52.2	853	1 NCAL_BOVIN	P11836 bos taurus
4	47.5	52.2	858	1 NCAL_RAT	P13596 rattus norv
5	47	51.6	935	1 ITH2_PIG	Q02668 sus scrofa
6	43	47.3	214	1 DEV5_MYXA	Q07766 myxococcus
7	43	47.3	596	1 MX67_SCHPO	Q5Y893 schizosacch
8	43	47.3	946	1 ITH2_HUMAN	P19823 homo sapien
9	42	46.2	108	1 IM13_CAEEL	Q45319 caenorhabdi
10	42	46.2	362	1 ARAR_BACSU	P66711 bacillus su
11	42	46.2	544	1 CH60_CHRPN	P16681 chlamydia p
12	42	46.2	1286	1 YKVS_YEAST	P28273 saccharomyc
13	42	46.2	1756	1 TR11_ECOLI	P14565 escherichia
14	42	46.2	1756	1 TR12_ECOLI	P22706 escherichia
15	42	46.2	2515	1 TUD_DROME	P25823 drosophila
16	41.5	45.6	1237	1 SM1B_HUMAN	Q8NDY3 homo sapien
17	41	45.1	150	1 YAO2_SCHPO	Q10081 schizosacch
18	41	45.1	344	1 ARGC_THRTN	Q8Y788 thermoaer
19	41	45.1	362	1 LEU3_SYNY3	P73960 synechocyst
20	41	45.1	479	1 PBP4_HABIN	P45161 h penicilli
21	41	45.1	524	1 UGTC_CAEEL	Q22181 caenorhabdi
22	41	45.1	539	1 G6PI_RALSO	Q8XY99 ralsomnia s
23	41	45.1	543	1 CH60_CHLMU	Q59322 chlamydia s
24	41	45.1	543	1 CH60_CHLIR	P17203 chlamydia t
25	41	45.1	629	1 YLIA_ECOLI	P75796 escherichia
26	41	45.1	641	1 TETS_LISMO	Q48791 listeria mo
27	41	45.1	646	1 TETS_LACIA	Q48712 lactococcus
28	41	45.1	682	1 VG50_BPMU5	Q05262 mycobacteri
29	41	45.1	946	1 INRI_MOUSE	Q61773 mus musculu
30	40.5	44.5	557	1 INRI_HUMAN	P17181 homo sapien
31	40	44.0	450	1 V50K_BYDP	P09516 barley yell
32	40	44.0	507	1 C4DE_DROME	Q46051 drosophila
33	40	44.0	544	1 CH61_CHICV	P15599 chlamydophi

34	40	44.0	565	1 CU23_SCHPO	Q94556 schizosacch
35	40	44.0	769	1 P1GR_RAT	P15083 rattus norv
36	39.5	43.4	536	1 CAR9_HUMAN	Q9H257 homo sapien
37	39.5	43.4	536	1 CAR9_RAT	Q9EP0 rattus norv
38	39	42.9	128	1 Y082_CAEEL	Q09237 caenorhabdi
39	39	42.9	155	1 P5AF_SYNY3	P29256 synechocyst
40	39	42.9	274	1 TRV1_ANOGA	P35035 anopheles g
41	39	42.9	362	1 LEU3_ANASP	Q8YX52 anabaena sp
42	39	42.9	402	1 CENC_SHEEP	P49453 ovis aries
43	39	42.9	430	1 SYN_STAMW	Q99U35 staphylococ
44	39	42.9	430	1 SYN_STAMW	Q8MP3 staphylococ
45	39	42.9	516	1 TRX2_DROME	Q9VNT5 drosophila

## ALIGNMENTS

RESULT 1  
ID ITH4\_HUMAN STANDARD; PRT; 930 AA.  
AC Q14624; Q15135; Q9P190; Q9U054;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Inter-alpha-trypsin inhibitor heavy chain H4 precursor (ITI heavy  
chain H4) (inter-alpha-inhibitor heavy chain 4) (inter-alpha-  
inhibitor family heavy chain-related protein) (IHRP) (Plasma  
DE kallikrein sensitive glycoprotein 120) (PK-120) (GPI20) (PRO1851)  
DE [Contains: GP57].  
GN ITH4 OR IHRP OR ITIH4 OR PK120.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NX NCBI\_TaxID=9606;  
RX [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND PARTIAL SEQUENCE.  
RC TISSUE=Liver;  
RX MEDLINE=95293915; PubMed=775381;  
RA Saguuchi K.-I., Tobe T., Hashimoto K., Sano Y., Nakano Y., Miura N.-H.,  
RA Tomita M.;  
RT "Cloning and characterization of cDNA for inter-alpha-trypsin  
inhibitor family heavy chain-related protein (IHRP), a novel human  
RT plasma glycoprotein.";  
RL J. Biochem. 117:14-18(1995).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND PARTIAL SEQUENCE.  
RC TISSUE=Liver;  
RX MEDLINE=95104473; PubMed=7805892;  
RA Nishimura H., Kakizaki I., Muta T., Sasaki N., Pu P.-X., Yamashita T.,  
RA Nagaawa S.;  
RT "cDNA and deduced amino acid sequence of human PK-120, a plasma  
RT kallikrein-sensitive glycoprotein.";  
RL FEBS Lett. 357:207-211(1995).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=96389955; PubMed=897089;  
RA Saguuchi K., Tobe T., Hashimoto K., Nagasaki Y., Oda E., Nakano Y.,  
RA Miura N.H., Tomita M.;  
RT "Isolation and characterization of the human inter-alpha-trypsin  
inhibitor family heavy chain-related protein (IHRP) gene (ITIH1).";  
RL J. Biochem. 119:898-905(1996).  
RN [4]  
RP SEQUENCE OF 271-930 FROM N.A. (ISOFORM 2).  
RC TISSUE=Fetal liver;  
RA Zhang C., Yu Y., Zhang S., Wei H., Zhang Y., Zhou G., Bi J., Liu M.,  
RA He F.;  
RT "Functional prediction of the coding sequences of 79 new genes deduced  
RT by analysis of cDNA clones from human fetal liver.";  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.  
RC TISSUE=plasma;  
RX MEDLINE=95332266; PubMed=7541790;

RA Choi-Miura N.-H., Sano Y., Oda E., Nakano Y., Tobe T., Yanagishita T.,  
 RA Taniyama M., Katagiri T., Tomita M.;  
 RT "Purification and characterization of a novel glycoprotein which has  
 RT significant homology to heavy chains of inter-alpha-trypsin inhibitor  
 RT family from human plasma.";  
 RL J. Biochem. 117:400-407(1995).  
 RN  
 RP VARIANT ASN-86.  
 RX PubMed:12147176;  
 RA Tozaki T., Choi-Miura N.-H., Taniyama M., Kurosawa M., Tomita M.;  
 RT "SNP analysis of the inter-alpha-trypsin inhibitor family heavy  
 RT chain-related protein (IHRP) gene by a fluorescence-adapted SSC  
 RT method.";  
 RL BMC Med. Genet. 3:6-6(2002).  
 CC -1- FUNCTION: MAY BE INVOLVED IN ACUTE PHASE REACTIONS.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q14624-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q14624-2; Sequence=VSP\_002761, VSP\_002762;  
 CC Note=No experimental confirmation available;  
 CC -1- TISSUE SPECIFICITY: Liver-specific.  
 CC -1- INDUCTION: LEVELS INCREASE SIGNIFICANTLY IN THE SERA OF PATIENTS  
 CC AFTER DIFFERENT SURGICAL TRAUMA.  
 CC -1- PTM: CLEAVED BY PLASMA KALLIKREIN TO YIELD 100- AND 35-KDA  
 CC FRAGMENTS, AND THE RESULTING 100-KDA FRAGMENT IS FURTHER CONVERTED  
 CC TO A 70-KDA FRAGMENT.  
 CC -1- PTM: APPEARS TO BE BOTH N- AND O-GLYCOSYLATED.  
 CC -1- SIMILARITY: BELONGS TO THE ITIH FAMILY.  
 CC -1- SIMILARITY: Contains 1 VWFA domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 DR EMBL; D38595; BAA07602.1; -;  
 DR EMBL; D38535; BAA07536.1; -;  
 DR EMBL; U43163; AAD05198.1; -;  
 DR EMBL; U42015; AAD05198.1; JOINED.  
 DR EMBL; U42016; AAD05198.1; JOINED.  
 DR EMBL; U43155; AAD05198.1; JOINED.  
 DR EMBL; U43156; AAD05198.1; JOINED.  
 DR EMBL; U43157; AAD05198.1; JOINED.  
 DR EMBL; U43158; AAD05198.1; JOINED.  
 DR EMBL; U43159; AAD05198.1; JOINED.  
 DR EMBL; U43160; AAD05198.1; JOINED.  
 DR EMBL; U43161; AAD05198.1; JOINED.  
 DR EMBL; U43162; AAD05198.1; JOINED.  
 DR EMBL; AF119856; AAF69610.1; -;  
 DR PIR; JX0368; JX0368.  
 DR Genem; HGNC:6169; ITIH4.  
 DR MIM; 600564; -;  
 DR GO; GO:0005211; F:Plasma glycoprotein; TAS.  
 DR InterPro; IPR006587; YIT.  
 DR InterPro; IPR002035; VWFA.  
 DR Pfam; PF00092; vwa; 1.  
 DR SMART; SM00609; YIT; 1.  
 DR SMART; SM00327; VWFA; 1.  
 DR PROSITE; PS50234; VWFA; 1.  
 KW Acute phase; Serine protease inhibitor; Repeat; Signal;  
 KW Multigene family; Glycoprotein; Alternative splicing; Polymorphism.  
 RT SIGNAL  
 FT CHAIN 1 28  
 FT CHAIN 29 661 70 kDa INTER-ALPHA-TRYPsin INHIBITOR  
 FT PROPEP 662 688 HEAVY CHAIN H4.  
 FT CHAIN 689 930 POTENTIALLY ACTIVE PEPTIDE.  
 FT CHAIN 930 35 kDa INTER-ALPHA-TRYPsin INHIBITOR  
 FT DOMAIN 272 432 HEAVY CHAIN H4.  
 FT VWFA.

FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 517 517 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 577 577 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 696 696 O-LINKED (GLCNAC) (PROBABLE).  
 FT CARBOHYD 701 701 O-LINKED (GLCNAC) (PROBABLE).  
 FT CARBOHYD 702 702 O-LINKED (GLCNAC) (PROBABLE).  
 FT DISULFID 747 747 MISSING (in isoform 2).  
 FT VARSPLIC 621 650 /FtId=VSP\_002761.  
 FT VARSPLIC 727 727 A -> ACPGSRAPAVPA (in isoform 2).  
 FT VARSPLIC 727 727 /FtId=VSP\_002762.  
 FT VARIANT 86 86 I -> N.  
 FT CONFLICT 85 85 /FtId=VAR\_013836.  
 FT CONFLICT 85 85 N -> I (IN REF. 2).  
 FT CONFLICT 85 85 N -> K (IN REF. 3).  
 FT CONFLICT 114 114 S -> N (IN REF. 1).  
 FT CONFLICT 207 207 N -> F (IN REF. 5).  
 FT CONFLICT 221 221 Q -> E (IN REF. 5).  
 FT CONFLICT 307 307 R -> V (IN REF. 5).  
 FT CONFLICT 322 322 W -> Y (IN REF. 5).  
 FT CONFLICT 816 817 ET -> OR (IN REF. 5).  
 SQ SEQUENCE 930 AA; 103358 MW; 0E1929065F4EB6A0 CRC64;  
 Query Match 100.0%; Score 91; DB 1; Length 930;  
 Best local similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Cy 1 KSPGQETVLDGNLIRY 18  
 Db 224 KSPGQETVLDGNLIRY 241  
 ID ITIH4\_PIG STANDARD; PRT; 921 AA.  
 AC P79263;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Inter-alpha-trypsin inhibitor heavy chain H4 precursor (ITI heavy  
 DE chain H4) (inter-alpha-inhibitor heavy chain 4) (inter-alpha-trypsin  
 DE inhibitor family heavy chain-related protein) (IHRP) (Major acute  
 DE phase protein) (MAP).  
 GN ITIH4 OR IHRP.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-36; 695-703 AND 710-722.  
 RC MEDLINE=96271024; PubMed=8830057;  
 RX Haehimoto K., Tobe T., Sumiya J.-I., Sano Y., Choi-Miura N.-H.,  
 RA Ozawa A., Yaeue H., Tomita M.;  
 RT "Primary structure of the pig homologue of human IHRP: inter-alpha-  
 RT trypsin inhibitor family heavy chain-related protein.";  
 RL J. Biochem. 119:577-584(1996).  
 RN [2]  
 RP PRELIMINARY SEQUENCE OF 267-556 FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=90371455; PubMed=1697703;  
 RA Buchanan T.G., Cabin D.E., Vickere S., Deutechman C.S., Delgado E.,  
 RA Sussman M.M., Bulkley G.B.;  
 RT "Molecular biology of circulatory shock. Part II. Expression of four  
 RT groups of hepatic genes is enhanced after resuscitation from  
 RT cardiogenic shock.";  
 RL Surgery 108:559-566(1990).  
 RN [3]  
 RP SEQUENCE OF 28-54 AND 223-240.  
 RC TISSUE=Serum;  
 RX MEDLINE=96013138; PubMed=7556597;  
 RA Gonzalez-Ramon N., Alava M.A., Sarsa J.A., Pineiro M., Escartin A.,

RA Garcia-Gil A., Lamprea F., Pineiro A.;  
 RT "The major acute phase serum protein in pigs is homologous to human  
 RT plasma kallikrein sensitive PK-120";  
 RL FEBS Lett. 371:227-230(1995).  
 CC -1- FUNCTION: MAY BE INVOLVED IN ACUTE PHASE REACTIONS.  
 CC -1- TISSUE SPECIFICITY: Liver-specific.  
 CC -1- INDUCTION: LEVELS INCREASE SIGNIFICANTLY AFTER CARDIOGENIC  
 CC SHOCK.  
 CC -1- PPM: APPEARS TO BE BOTH N- AND O-GLYCOSYLATED (BY SIMILARITY).  
 CC -1- PPM: CLEAVED BY PLASMA KALLIKREIN TO YIELD 55- AND 25-KDA  
 CC FRAGMENTS.  
 CC -1- SIMILARITY: BELONGS TO THE ITIH FAMILY.  
 CC -1- CAUTION: REF.2 SEQUENCE IS INCORRECT DUE TO FRAMESHIFTS AND OTHER  
 CC SEQUENCING ERRORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U43164; AAD00024.1; -;  
 CC EMBL; S82800; AAB46821.1; -;  
 CC EMBL; M29507; -; NOT\_ANNOTATED\_CDS.  
 CC PIR; JG4625; JG4625.  
 CC InterPro; IPR005587; VIT.  
 CC InterPro; IPR020351; VWF\_A.  
 CC Pfam; PF00092; VWA; 1.  
 CC SMART; SM00609; VIT; 1.  
 CC SMART; SM00327; VWA; 1.  
 CC PROSITE; PS50234; VWF\_A; 1.  
 CC Serine protease inhibitor; Repeat; Signal; Multigene family;  
 KM GLYCOPROTEIN.  
 FT SIGNAL 1 27  
 FT CHAIN 28 921  
 FT  
 FT INTER-ALPHA-TRYPsin INHIBITOR HEAVY CHAIN  
 FT H4.  
 FT DOMAIN 270 428 VWF\_A.  
 FT CARBOHYD 80 80 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 205 205 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 242 242 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 513 513 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 577 577 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 49 50 HT -> SK (IN REF. 3).  
 FT CONFLICT 703 703 D -> H (IN REF. 1; AA SEQUENCE).  
 SQ SEQUENCE 921 AA; 102146 MW; E2BF9525DE8D07C CRC64;  
 Query March 94.5%; Score 86; DB 1; Length 921;  
 Best Local Similarity 88.9%; Pred. No. 1.1e-06;  
 Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 KSPROQETVLGDLNLIIRY 18  
 Db 222 KSPROQETVLGDLNLIIRY 239  
 RESULT 3  
 NCAL BOVIN STANDARD; PRT; 853 AA.  
 ID NCAL BOVIN STANDARD; PRT; 853 AA.  
 AC P31836;  
 DT 01-JUN-1993 (Rel. 26, Created)  
 DT 01-JUN-1993 (Rel. 26, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Neural cell1 adhesion molecule 1, 140 kDa isoform precursor (N-CAM 140)  
 DE (NCAM-140).  
 GN NCAM1 OR NCAM.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxId=9913;

EN [1]  
 RC SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Brain cortex;  
 RX MEDLINE=69378239; PubMed=2776887;  
 RA Lipkin V.M., Khramtsov N.V., Andreeva S.G., Moshayakov M.V.,  
 RA Petukhova G.V., Rakitina T.V., Peshchenko E.A., Ishchenko K.A.,  
 RA Mirzoeva S.F., Chernova M.N., Dranytsyna S.M.;  
 RT "Calciumulin-independent bovine brain adenylyl cyclase. Amino acid  
 RT sequence and nucleotide sequence of the corresponding cDNA.";  
 RL FEBS Lett. 254:69-73(1989).  
 RN [2]  
 RP SEQUENCE OF 20-36.  
 RX MEDLINE=66140120; PubMed=3512556;  
 RA Rougon G., Marshak D.R.;  
 RT "Structural and immunological characterization of the amino-terminal  
 RT domain of mammalian neural cell adhesion molecules.";  
 RL J. Biol. Chem. 261:3396-3401(1986).  
 RN [3]  
 RP IDENTIFICATION AS N-CAM.  
 RX MEDLINE=92111748; PubMed=1765159;  
 RA Premont R.T.;  
 RT "A bovine brain cDNA purported to encode calmodulin-insensitive  
 RT adenylyl cyclase has extensive identity with neural cell adhesion  
 RT molecules (N-CAMs).";  
 RL FEBS Lett. 295:230-231(1991).  
 CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN  
 CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF  
 CC NEURITES, ETC.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=1;  
 CC Comment=A number of isoforms are produced;  
 CC Name=1;  
 CC IsoId=P31836-1; Sequence=displayed;  
 CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.  
 CC -1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.  
 CC -1- SIMILARITY: Contains 2 fibronectin type III domains.  
 CC -1- CAUTION: WAS ORIGINALLY (REF.1) THOUGHT TO BE A CALMODULIN-  
 CC INDEPENDENT ADENYLYL CYCLASE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X16451; CAA34470.1; -;  
 CC PIR; A32976; IIBONC.  
 DR HSSP; P40189; IBOU.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003598; Ig\_C2.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR Pfam; PF00041; fn3; 2.  
 DR Pfam; PF00047; Ig; 5.  
 DR SMART; SM00060; FN3; 2.  
 DR SMART; SM00408; IGC2; 5.  
 DR PROSITE; PS50835; IG LIKE; 5.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat;  
 KW Immunoglobulin domain; Alternative splicing; Signal.  
 KM SIGNAL 1 19  
 FT CHAIN 20 853  
 FT  
 FT NEURAL CELL ADHESION MOLECULE 1, 140 kDa  
 FT ISOFORM.  
 FT DOMAIN 20 719  
 FT TRANSMEM 720 737  
 FT DOMAIN 738 853  
 FT DOMAIN 20 111  
 FT DOMAIN 116 205  
 FT DOMAIN 212 300  
 FT DOMAIN 307 412  
 FT DOMAIN 415 500  
 FT IG-LIKE C2-TYPE 1.  
 FT IG-LIKE C2-TYPE 2.  
 FT IG-LIKE C2-TYPE 3.  
 FT IG-LIKE C2-TYPE 4.  
 FT IG-LIKE C2-TYPE 5.

```

FT DOMAIN 527 604 FIBRONECTIN TYPE-III 1.
FT DOMAIN 633 700 FIBRONECTIN TYPE-III 2.
FT DOMAIN 152 156 HEPARIN-BINDING (POTENTIAL).
FT DOMAIN 161 165 HEPARIN-BINDING (POTENTIAL).
FT DISULFID 41 96 BY SIMILARITY.
FT DISULFID 139 189 BY SIMILARITY.
FT DISULFID 235 286 BY SIMILARITY.
FT DISULFID 328 394 BY SIMILARITY.
FT DISULFID 435 488 BY SIMILARITY.
FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 314 314 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 346 346 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 458 458 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 487 487 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 853 AA; 93893 MW; E12FD49231A7A368 CRC64;

Query Match 52.2%; Score 47.5; DB 1; Length 853;
Best Local Similarity 60.0%; Pred. No. 4.1;
Matches 9; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

OY 3 PEOQETVLDGNLIR 17
Db 358 PEOQET-LDGMHVR 371

RESULT 4
NCAL RAT STANDARD; PRT; 858 AA.
AC P1356;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Neural cell adhesion molecule 1, 140 kDa isoform precursor (N-CAM 140)
DE (NCAM-140).
GN NCAM1 OR NCAM.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=88059265; PubMed=3680385;
RA Small S.U., Shull G.E., Santoni M.-J., Akeson R.;
RT "Identification of a cDNA clone that contains the complete coding
RT sequence for a 140-kD rat NCAM polypeptide.";
RL J. Cell Biol. 105:2335-2345(1987).
RN [2]
RP SEQUENCE OF 355-364 FROM N.A.
RX MEDLINE=90166485; PubMed=2483093;
RA Small S.U., Haines S.L., Akeson R.A.;
RT "Polypeptide variation in an N-CAM extracellular immunoglobulin-like
RT fold is developmentally regulated through alternative splicing.";
RL Neuron 1:1007-1017(1988).
CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE INVOLVED IN
CC NEURON-NEURON ADHESION, NEURITE FASCICULATION, OUTGROWTH OF
CC NEURITES, ETC.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=1;
CC Comment=A number of isoforms are produced;
CC Name=1;
CC IsoId=P1356-1; Sequence=Displayed;
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 2 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

---

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X06564; CAA23809.1; -.
DR EMBL; M32611; AAA41679.1; -.
DR PIR; S00846; IJRTNC.
DR PDB; 1BPF; 27-OCT-00.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; fn3; 2.
DR Pfam; PF00047; Ig; 5.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00408; IgC2; 5.
DR PROSITE; PS50835; IG_LIKE; 5.
KW Cell adhesion; Glycoprotein; Transmembrane; Repeat;
KW Immunoglobulin domain; Alternative splicing; Signal; Heparin-binding;
KW 3D-structure.
FT SIGNAL 1 19
FT CHAIN 20 858
FT FT
FT DOMAIN 20 721
FT TRANSMEM 722 739
FT DOMAIN 740 858
FT DOMAIN 20 111
FT DOMAIN 116 205
FT DOMAIN 212 302
FT DOMAIN 309 414
FT DOMAIN 417 502
FT DOMAIN 514 615
FT DOMAIN 616 712
FT DOMAIN 152 156
FT DOMAIN 161 165
FT DISULFID 41 96
FT DISULFID 139 189
FT DISULFID 235 288
FT DISULFID 330 396
FT DISULFID 437 490
FT CARBOHYD 222 222
FT CARBOHYD 316 316
FT CARBOHYD 348 348
FT CARBOHYD 434 434
FT CARBOHYD 460 460
FT CARBOHYD 489 489
SQ SEQUENCE 858 AA; 94658 MW; E1A064A4EA050B6 CRC64;

Query Match 52.2%; Score 47.5; DB 1; Length 858;
Best Local Similarity 60.0%; Pred. No. 4.2;
Matches 9; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

OY 3 PEOQETVLDGNLIR 17
Db 360 PEOQET-LDGMHVR 373

RESULT 5
ITH2_PIG STANDARD; PRT; 935 AA.
AC O02668;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Inter-alpha-trypsin inhibitor heavy chain H2 precursor (ITI heavy
DE chain H2) (Inter-alpha-inhibitor heavy chain 2).
GN ITH2.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
OX NCBI_TaxID=9623;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=liver;

```

```

RA Gebhard W.;
CC Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY ACT AS A CARRIER OF HYALURONAN IN SERUM OR AS A
CC BINDING PROTEIN BETWEEN HYALURONAN AND OTHER MATRIX PROTEIN,
CC INCLUDING THOSE ON CELL SURFACES IN TISSUES TO REGULATE THE
CC LOCALIZATION, SYNTHESIS AND DEGRADATION OF HYALURONAN WHICH ARE
CC ESSENTIAL TO CELLS UNDERGOING BIOLOGICAL PROCESSES (BY
CC SIMILARITY).
CC -1- SUBUNIT: 1-ALPHA-1 PLASMA PROTEASE INHIBITORS ARE ASSEMBLED FROM
CC ONE OR TWO HEAVY CHAINS (H1, H2 OR H3) AND ONE LIGHT CHAIN,
CC BIKUNIN. INTER-ALPHA-INHIBITOR (1-ALPHA-1) IS COMPOSED OF H1, H2
CC AND BIKUNIN. INTER-ALPHA-LIKE INHIBITOR (1-ALPHA-LI) OF H2 AND
CC BIKUNIN. AND PRE-ALPHA-INHIBITOR (P-ALPHA-1) OF H3 AND BIKUNIN.
CC -1- PTM: HEAVY CHAINS ARE INTERLINKED WITH BIKUNIN VIA A CHONDROITIN
CC 4-SULFATE BRIDGE TO THEIR C-TERMINAL ASPARTATE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ITIH FAMILY.
CC -1- SIMILARITY: Contains 1 VWFA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y11545; CAAT2308.1; -.
DR InterPro; IPR006587; VIT.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; VWF; 1.
DR SMART; SM00609; VIT; 1.
DR SMART; SM00327; VWF; 1.
DR PROSITE; PS50234; VWFA; 1.
KW Serine protease inhibitor; Repeat; Signal; Multigene family;
KM Glycoprotein.
FT SIGNAL 1 18 POTENTIAL.
FT PROPEP 19 53 BY SIMILARITY.
FT CHAIN 54 691 INTER-ALPHA-TRYPSIN INHIBITOR HEAVY CHAIN
FT PROPEP 692 935 BY SIMILARITY.
FT DOMAIN 297 457 VWFA.
FT CARBOHYD 107 107 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 434 434 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT BINDING 691 691 CHONDROITIN 4-SULFATE, CROSS-LINK SITE
FT (BY SIMILARITY).
SQ SEQUENCE 935 AA; 104621 MW; 88DAAC1BED367CC CRC64;

Query Match 51.6%; Score 47; DB 1; Length 935;
Best Local Similarity 44.4%; Pred. No. 5.6;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 KSPQOETVLDGNLIIRY 18
DB 249 KCPSCSETAVDGLVVMY 266

RESULT 6
DEVS_MYXXA STANDARD; PRT; 214 AA.
AC 007766;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Fruiting body developmental protein S.
GN DEVS.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;
OC Cythobacteriineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxId=34;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DK1622;

```

```

RX MEDLINE=94042921; PubMed=7693658;
RA Thoeny-Meyer L., Kaiser D.;
RT "DEVS", an autoregulated and essential genetic locus for fruiting
RT body development in Myxococcus xanthus. ";
RL J. Bacteriol. 175:7450-7462 (1993).
CC -1- FUNCTION: HAS A ROLE IN FRUITING BODY DEVELOPMENT, SPOULATION
CC AND AGGREGATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L19029; AAA16135.1; -.
DR PIR; C49941; C49941.
KW Fruiting body; Sporulation.
SQ SEQUENCE 214 AA; 23302 MW; 748EA24915A86675 CRC64;

Query Match 47.3%; Score 43; DB 1; Length 214;
Best Local Similarity 40.0%; Pred. No. 5.1;
Matches 6; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 SPEQOETVLDGNLIIR 16
DB 88 APDQOQVLDARLVV 102

RESULT 7
MX67 SCHPO STANDARD; PRT; 596 AA.
AC 09Y8G3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE mRNA export factor mex67.
GN MEX67 OR SPBC1921.03C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxId=4896;
RN [1]
RP SEQUENCE FROM N.A. AND FUNCTION.
RC STRAIN=972;
RX MEDLINE=20528620; PubMed=11073978;
RA Yoon J.H., Love D.C., Gubhakurta A., Hanover J.A., Dhar R.;
RT "Mex67p of Schizosaccharomycetes pombe interacts with Raelp1 in mediating
RT mRNA export. ";
RL Mol. Cell. Biol. 20:8767-8782 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream W.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gantles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean U.,
RA Mooney P., Moutle S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Wolckeste G., Aert R., Robben J., Gymnopoulos B.,
RA Weltegens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,

```

```

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.-J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.,
RA "The genome sequence of Schizosaccharomyces pombe.";
RA Nature 415:871-880(2002).
RL
CC -1- FUNCTION: Involved in the export of mRNA from the nucleus to the
CC cytoplasm.
CC -1- SUBUNIT: Interacts with rael.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE NXF FAMILY.
CC -1- SIMILARITY: Contains 1 NTF2 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AF055036; AAD43831.1; -.
DR EMBL, AL122033; CAB58969.1; -.
DR PIR, T43683; T43683.
DR GenDB, SPombe; SPBC1921.03c; -.
DR InterPro; IPR002075; NTF2.
DR InterPro; IPR005637; TAP_C.
DR Pfam; PF03943; TAP_C; 1.
DR PROSITE; PS50177; NTF2_DOMAIN; 1.
DR Transport; mRNA transport; Nuclear protein.
DR FT DOMAIN 338 499 NTF2.
DR SEQUENCE 596 AA; 66534 MW; B84345B8C0D3D83 CRC64;
SQ
Query Match 47.3%; Score 43; DB 1; Length 596;
Best Local Similarity 62.5%; Pred. No. 16;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
Oy 1 KSPGQETVLDPGNLI 16
Db 266 KLPELQELVLVGNPIV 281

```

```

RA Benarous R., Okubo I., Kurachi S., Kurachi K., Martin J.P.;
RA "Isolation and characterization of cDNAs encoding the heavy chain of
RA human inter-alpha-trypsin inhibitor (I alpha TI): unambiguous
RA evidence for multipolypeptide chain structure of I alpha TI.";
RA Proc. Natl. Acad. Sci. U.S.A. 84:8272-8276(1987).
RL
RN [3]
RP SEQUENCE OF 384-766 FROM N.A.
RX MEDLINE=89076497; PubMed=2462430;
RA Salter J.-P., Diarra-Mehrpour M., Seeboue R., Bourguignon J.,
RA Martin J.P.;
RA "Human inter-alpha-trypsin inhibitor. Isolation and characterization
RA of heavy (H) chain cDNA clones coding for a 383 amino-acid sequence
RA of the H chain.";
RA Biol. Chem. Hoppe-Seyler 369:15-18(1988).
RL
RN [4]
RP PARTIAL SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=88024442; PubMed=3663330;
RA Schreitmuller T., Hochstrasser K., Reisinger P.W.M., Wachter E.,
RA Gebhard W.;
RA "cDNA cloning of human inter-alpha-trypsin inhibitor discloses three
RA different proteins ";
RA Biol. Chem. Hoppe-Seyler 368:963-970(1987).
RL
RN [5]
RP SEQUENCE OF 55-74, 116-127, 224-246, 295-307 AND 365-385.
RX MEDLINE=89380192; PubMed=2476436;
RA Enghild J.J., Thøgersen I.B., Pizzo S.V., Salvesen G.;
RA "Analysis of inter-alpha-trypsin inhibitor and a novel trypsin
RA inhibitor, pre-alpha-trypsin inhibitor, from human plasma.
RA Polypeptide chain stoichiometry and assembly by glycan.";
RA J. Biol. Chem. 264:15975-15981(1989).
RN
RN [6]
RP SEQUENCE OF 55-64.
RC TISSUE=Plasma;
RX MEDLINE=93039735; PubMed=1384548;
RA Malki N., Balduyck M., Maes P., Capon C., Mizon K.K.,
RA Tartar A., Fournet B., Mizon J.;
RA "The heavy chains of human plasma inter-alpha-trypsin inhibitor: their
RA isolation, their identification by electrophoresis and partial
RA sequencing. Differential reactivity with concanavalin A.";
RA Biol. Chem. Hoppe-Seyler 373:1009-1018(1992).
RL
RN [7]
RP SEQUENCE OF 55-64 AND 681-702, CROSS-LINK STRUCTURE, AND
RP CARBOHYDRATE-LINKAGE SITES THR-691.
RX MEDLINE=93232026; PubMed=7682553;
RA Enghild J.J., Salvesen G., Thøgersen I.B., Valnickova Z.,
RA Pizzo S.V., Hefta S.A.;
RA "Presence of the protein-glycosaminoglycan-protein covalent cross-link
RA in the inter-alpha-inhibitor-related proteinase inhibitor heavy chain
RA 2/bikunin.";
RA J. Biol. Chem. 268:8711-8716(1993).
RN
RN [8]
RP SEQUENCE OF 67-101, AND HYALURONAN BINDING.
RC TISSUE=Serum;
RX MEDLINE=94075371; PubMed=7504674;
RA Huang L., Yoneda M., Kimata K.;
RA "A serum-derived hyaluronan-associated protein (SHAP) is the heavy
RA chain of the inter alpha-trypsin inhibitor.";
RA J. Biol. Chem. 268:26725-26730(1993).
RL
RN [9]
RP SEQUENCE OF 699-702, AND COVALENT LINKAGE WITH CHONDROITIN SULFATE.
RC TISSUE=Plasma;
RX MEDLINE=94228087; PubMed=7513643;
RA Morelle W., Capon C., Balduyck M., Sautiere P., Kousch M.,
RA Michalek C., Fournet B., Mizon J.;
RA "Chondroitin sulphate covalently cross-links the three polypeptide
RA chains of inter-alpha-trypsin inhibitor.";
RA Eur. J. Biochem. 221:881-888(1994).
RN
RN [10]
RP CARBOHYDRATE-LINKAGE SITES, AND MASS SPECTROMETRY.
RX MEDLINE=98343966; PubMed=9677337;
RA Flahaut C., Capon C., Balduyck M., Ricart G., Sautiere P., Mizon J.;
RA "Glycosylation pattern of human inter-alpha-inhibitor heavy chains.";

```



RL Biochem. J. 333:749-756(1998).  
 RN (11)  
 RP CARBOHYDRATE-LINKAGE SITES, AND DISULFIDE BONDS.  
 RX MEDLINE=98087700; PubMed=9425062;  
 RA Olsen E.H.N., Rønbek-Nielsen H., Thøgersen I.B., Roepstorff P.,  
 RA Engild J.J.;  
 RT "Posttranslational modifications of human inter-alpha-inhibitor:  
 RT identification of glycans and disulfide bridges in heavy chains 1 and  
 RT 2";  
 RL Biochemistry 37:408-416(1998).  
 CC -1- FUNCTION: MAY ACT AS A CARRIER OF HYALURONAN IN SERUM OR AS A  
 CC BINDING PROTEIN BETWEEN HYALURONAN AND OTHER MATRIX PROTEIN,  
 CC INCLUDING THOSE ON CELL SURFACES IN TISSUES TO REGULATE THE  
 CC LOCALIZATION, SYNTHESIS AND DEGRADATION OF HYALURONAN WHICH ARE  
 CC ESSENTIAL TO CELLS UNDERGOING BIOLOGICAL PROCESSES.  
 CC -1- SUBUNIT: I-ALPHA-1 PLASMA PROTEASE INHIBITORS ARE ASSEMBLED FROM  
 CC ONE OR TWO HEAVY CHAINS (H1, H2 OR H3) AND ONE LIGHT CHAIN,  
 CC BIKUNIN. INTER-ALPHA-INHIBITOR (I-ALPHA-I) IS COMPOSED OF H1, H2  
 CC AND BIKUNIN, INTER-ALPHA-LIKE INHIBITOR (I-ALPHA-LI) OF H2 AND  
 CC BIKUNIN, AND PRE-ALPHA-INHIBITOR (P-ALPHA-I) OF H3 AND BIKUNIN.  
 CC -1- PFM: HEAVY CHAINS ARE INTERLINKED WITH BIKUNIN VIA A CHONDROITIN  
 CC 4-SULFATE BRIDGE TO THEIR C-TERMINAL ASPARTATE.  
 CC -1- MASS SPECTROMETRY: MW=76508; METHOD=MALDI; RANGE=55-702.  
 CC -1- SIMILARITY: BELONGS TO THE ITIH FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X07173; CA30160.1; ALT\_SEQ.  
 DR EMBL: M18193; AAA60558.1; -.  
 DR EMBL: M33033; AAA59195.1; -.  
 DR PIR: S00346; IYH2.  
 DR GYCSU000001; P19823; -.  
 DR Genew: HGNC:6167; ITIH2.  
 DR MIM: 146640; -.  
 DR GO: GO:0005211; F: Plasma glycoprotein; TAS.  
 DR InterPro: IPR006587; VWFA.  
 DR InterPro: IPR002035; VWFA\_A.  
 DR Pfam: PF00092; Vwa; 1.  
 DR SMART: SM00609; VIT; 1.  
 DR SMART: SM00327; VMA; 1.  
 DR PROSITE: PS50234; VWFA; 1.  
 KM Serine protease inhibitor; Repeat: Signal; Multigene family;  
 KM Gamma-carboxyglutamic acid; Glycoprotein.  
 FT SIGNAL 1 18  
 FT PROPEP 19 54  
 FT CHAIN 55 702  
 FT  
 FT PROPEP 703 946  
 FT DOMAIN 308 468  
 FT DISULFID 261 264  
 FT DISULFID 650 651  
 FT CARBOHYD 118 118  
 FT  
 FT CARBOHYD 666 666  
 FT  
 FT CARBOHYD 671 671  
 FT CARBOHYD 673 673  
 FT CARBOHYD 675 675  
 FT CARBOHYD 691 691  
 FT  
 FT MOD RES 282 282  
 FT MOD RES 283 283  
 FT BINDING 702 702  
 FT CONFLICT 374 374

FT CONFLICT 674 674 P -> A (IN REF. 2 AND 3).  
 FT CONFLICT 705 705 F -> S (IN REF. 2 AND 3).  
 FT CONFLICT 729 729 N -> D (IN REF. 2 AND 3).  
 FT CONFLICT 731 731 V -> A (IN REF. 2 AND 3).  
 SQ SEQUENCE 946 AA; 106436 MW; 1478CFB8F3BA776 CRC64;  
 Query Match 47.3%; Score 43; DB 1; Length 946;  
 Best Local Similarity 43.8%; Pred. No. 28;  
 Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
 Oy 3 PEOQETVLDGNIIRY 18  
 Db 262 PSCEIYVDELVLY 277  
 RESULT 9  
 ID TIM3\_CABEL STANDARD; PRT; 108 AA.  
 AC 045319;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Mitochondrial Import inner membrane translocase subunit Tim13.  
 GN TIM-13 OR TIM13 OR DY3.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bauer M.F., Brunner M., Hofmann S.;  
 RT "Cloning and mapping of the Tim10/DDP gene family encoding small zinc  
 RT finger proteins involved in mitochondrial carrier import.";  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Lennard N.;  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Likely to be involved in the import and insertion of  
 CC hydrophobic membrane proteins into the mitochondrial inner  
 CC membrane (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane (By  
 CC similarity).  
 CC -1- Similarity: Belongs to the Tim8/Tim10 family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF144704; AAD39955.1; -.  
 DR EMBL: Z96047; CAB09410.1; -.  
 DR PIR: T20389; T20389.  
 DR WormPep: DY3.1; CE15745.  
 DR InterPro: IPR004217; Znf\_Tim10/DDP.  
 DR Pfam: PF02953; zf-Tim10 DDP; 1.  
 KM Transport; Protein transport; Translocation; Mitochondrion;  
 KM Inner membrane.  
 SQ SEQUENCE 108 AA; 11685 MW; 7CF2DA7BC6B229AB CRC64;  
 Query Match 46.2%; Score 42; DB 1; Length 108;  
 Best Local Similarity 72.7%; Pred. No. 35;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Oy 2 SPEQOETVLDG 12  
 Db 14 SPEQOEVVISG 24

```

RESULT 10
ARAR_BACSU STANDARD; PRT; 362 AA.
ID ARAR_BACSU STANDARD; PRT; 362 AA.
AC P96711;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Arabinose metabolism transcriptional repressor.
OS ARAR OR ARAC.
NS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
CX NCBI_TaxID=1423;
RN [1]
RP STRAIN=168;
RC MEDLINE=97197523; PubMed=9045819;
RA Sa-Nogueira I.M.G., Mota L.J.;
RT "Negative regulation of L-arabinose metabolism in Bacillus subtilis:
RL J. Bacteriol. 179:1598-1608(1997).
[2]
RP STRAIN=168;
RC MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Berteiro M.G., Bessières P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brigelli S.C., Bron S.,
RA Brouillet S., Brunsch C.V., Caldwell B., Capuano N.J., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kaasara Y., Klaerr-Blanchard M., Klein C.,
RA Kodayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Patro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Priesean E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Seliguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpeira P., Tognoni A.,
RA Tosato V., Uchlyama S., Vandenbol M., Vannier F., Vasarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zanchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RL subtilis."
RL Nature 390:249-256(1997).
[3]
RP STRAIN=168;
RC MEDLINE=99348379; PubMed=10417639;
RA Mota L.J., Tavares P., Sa-Nogueira I.M.G.;
RT "Mode of action of Arar, the key regulator of L-arabinose metabolism
RL in Bacillus subtilis."
RL Mol. Microbiol. 33:476-489(1999).
[4]
RP STRAIN=168;
RC MEDLINE=21311744; PubMed=11418559;
RA Mota L.J., Sarmiento L.M., Sa-Nogueira I.M.G.;
RT "Control of the arabinose regulon in Bacillus subtilis by Arar in
RL vivo: crucial roles of operators, cooperativity, and DNA looping."
RL J. Bacteriol. 183:4190-4201(2001).
-1- FUNCTION: Transcriptional repressor of the arabinose utilization
CC genes. Also regulates its own expression. Binds to two sequences
CC within the promoters of the araBDLMWPQ-abfA operon and the araB
CC

```

```

CC gene, and to one sequence in the arar promoter.
CC -1- ENZYME REGULATION: Binding to DNA is inhibited by L-arabinose.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE GNTR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC
CC -----
CC EMBL; X98354; CA66999.1; -.
CC EMBL; Z99121; CAB15402.1; ALT_INIT.
CC HSSP; P15039; IDBO.
CC Subtilist; BG11913; arar.
CC InterPro; IPR000524; HTH_GntR.
CC InterPro; IPR001761; PeriplabP/LacI.
CC Pfam; PF00392; gntR; 1.
CC Pfam; PF00532; Peripla_BP_Like; 1.
CC PRINTS; PR00035; HTHGNT.
CC SMART; SM00345; HTH_GNT.
CC PROSITE; PS00043; HTH_GNT.
CC TRANSIT; T10000; H-T-H MOTIF (POTENTIAL).
CC DNA BIND 30 49
CC SEQUENCE 362 AA; 40487 MW; 679DC8FE69B421 CRC64;
CC
CC Query Match 46.2%; Score 42; DB 1; Length 362;
CC Best Local Similarity 41.2%; Pred. No. 14;
CC Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
CC
CC 1 KSPQCEVTDGNLIIR 17
CC Db 338 KKPQEDVTFPEELIR 354
CC
CC RESULT 11
CC CH60_CHLPN STANDARD; PRT; 544 AA.
CC ID CH60_CHLPN STANDARD; PRT; 544 AA.
CC AC P31681; Q9J079;
CC DT 01-JUL-1993 (Rel. 26, Created)
CC DT 30-MAY-2000 (Rel. 39, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE 60 kDa chaperonin (Protein Cpn60) (groEL protein).
CC GN GROEL OR GROEL OR MOXA OR CPN0134 OR CP0638.
CC OS Chlamydia pneumoniae (Chlamydia pneumoniae).
CC OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
CC NCBI_TaxID=83558;
CC [1]
CC RP STRAIN=AR39;
CC RC MEDLINE=92040169; PubMed=1682266;
CC RA Kikuta L.C., Puolakainen M., Kuo C.C., Campbell L.A.;
CC RT "Isolation and sequence analysis of the Chlamydia pneumoniae GroE
CC operon."
CC Infect. Immun. 59:4665-4669(1991).
CC [2]
CC RP STRAIN=CWL029;
CC RC MEDLINE=99206606; PubMed=10192388;
CC RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
CC RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
CC RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."
CC Nat. Genet. 21:385-389(1999).
CC [3]
CC RP STRAIN=AR39;
CC RC MEDLINE=20150255; PubMed=10684935;
CC RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
CC RA White O., Hickey E.K., Peterson J., Ueberlack T., Berry K., Baas S.,
CC Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
CC

```

RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia  
 RT pneumoniae AR3.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CML029 from USA.";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 CC -1- FUNCTION: Prevents misfolding and promotes the refolding and  
 CC proper assembly of unfolded polypeptides generated under stress  
 CC conditions (By similarity).  
 CC -1- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of  
 CC 7 subunits (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- INDUCTION: By stress.  
 CC -1- SIMILARITY: Belongs to the chaperonin (HSP60) family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
 CC -----  
 DR EMBL: M69217; AAA23126.1; -;  
 DR EMBL: X60068; CAA42673.1; -;  
 DR EMBL: AEO01500; AAD18287.1; -;  
 DR EMBL: AEO02221; AAF38453.1; -;  
 DR EMBL: AP002545; BAA98344.1; -;  
 DR PIR: B81556; B81556;  
 DR PIR: F86507; F86507;  
 DR PIR: S19023; S19023;  
 DR HSSP: P06139; 1GRU.  
 DR PHCT-2DPAGE; P31681; -;  
 DR TIGR: CP0638; -;  
 DR HAMAP: MF\_00600; -; 1.  
 DR InterPro: IPR001844; Chaperonin\_Cpn60.  
 DR InterPro: IPR002423; Cpn60\_TCE-1.  
 DR Pfam: PF00118; Cpn60\_TCE1.1.  
 DR PRINTS: PR00298; CHAPERONIN60.  
 DR PROSITE: PS00304; TCOMPLEXTCP1.  
 DR PROSITE: PS00296; CHAPERONIN60; 1.  
 KM Chaperone; ATP-binding; Complete proteome.  
 FT CONFLICT 510 510 A -> R (IN REF. 1).  
 SQ SEQUENCE 544 AA; 58203 MW; 632BE406F5112CA4F CRC64;  
 Query Match 46.2%; Score 42; DB 1; Length 544;  
 Best Local Similarity 52.9%; Pred. No. 22;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
 QY 2 SPEQOETVLDGNLTIRY 18  
 Db 207 NPETOECVLEDAIILTY 223  
 ID YKVS\_YEAST STANDARD; PRT; 1286 AA.  
 ID YKVS\_YEAST  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 140.4 kDa protein in URA1-DOA1 intergenic region.  
 GN YKJ215C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C;  
 RX MEDLINE=95028164; PubMed=7941750;  
 RA Tzeremla M., Horaitis O., Alexandraki D.;  
 RT "The complete sequencing of a 24.6 kb segment of yeast chromosome XI  
 RT identified the known loci URA1, SAG1 and TRP3, and revealed 6 new  
 RT open reading frames including homologues to the threonine  
 RT dehydratase, membrane transporters, hydantoinsases and the  
 RT phospholipase A2-activating protein.";  
 RL Yeast 10:663-679(1994).  
 RN [2]  
 RP SEQUENCE OF 1-1182 FROM N.A.  
 RA Pohl T.M., Pohl F.M.;  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1003-1285 FROM N.A.  
 RC STRAIN=ATCC 28383 / FL100.  
 RX MEDLINE=92380485; PubMed=1511880;  
 RA Roy A.;  
 RT "Nucleotide sequence of the URA1 gene of Saccharomyces cerevisiae.";  
 RL Gene 118:149-150(1992).  
 CC -1- SIMILARITY: BELONGS TO THE OXOPROLINASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
 CC -----  
 DR EMBL: X75951; CAA53558.1; -;  
 DR EMBL: Z28215; CAA82060.1; -;  
 DR EMBL: X59371; CAA42015.1; -;  
 DR EMBL: M83295; AAA34567.1; -;  
 DR PIR: S38058; S38058.  
 DR SCD; S0001698; YKJ215C.  
 DR InterPro: IPR002821; Hydantoinsase\_A.  
 DR InterPro: IPR003692; Hydantoinsase\_B.  
 DR Pfam: PF05378; Hydant\_A\_N\_1.  
 DR Pfam: PF01968; Hydantoinsase\_A; 1.  
 DR Pfam: PF02538; Hydantoinsase\_B; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 1286 AA; 140427 MW; 9A0B8C609B5D6FF6 CRC64;  
 Query Match 46.2%; Score 42; DB 1; Length 1286;  
 Best Local Similarity 58.3%; Pred. No. 58;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 SPEQOETVLDGN 13  
 Db 156 SPWEGEGILGN 167  
 ID TRII\_ECOLI STANDARD; PRT; 1756 AA.  
 ID TRII\_ECOLI  
 AC P14555; O51811;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Trp1 protein (DNA helicase I) (EC 3.6.1.-) [Contains: Trp1\* protein].  
 GN TRAI.  
 OS Escherichia coli.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

```

RX MEDLINE=90299847; PubMed=2163400;
RA Bradshaw H.D. Jr., Traxler B.A., Minkley E.G. Jr., Neeter E.W.,
RA Gordon M.P.;
RT "Nucleotide sequence of the traI (helicase I) gene from the sex
RT factor F.";
RL J. Bacteriol. 172:4127-4131 (1990).
RN [2]
RP MEDLINE=94359430; PubMed=7915817;
RX Frost L.S., Ippen-Inler K., Skurray R.A.;
RA "Analysis of the sequence and gene products of the transfer region of
RT the F sex factor.";
RL Microbiol. Rev. 58:162-210 (1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / CR63;
RA Shimizu H., Saitoh Y., Suda Y., Uehara K., Sampei G., Mizobuchi K.;
RT "Complete nucleotide sequence of the F plasmid: its implications for
RT organization and diversification of plasmid genomes.";
RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-150 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90034191; PubMed=2680768;
RA Jatajakumari M.B., Manning P.A.;
RT "Nucleotide sequence of the traD region in the Escherichia coli F sex
RT factor.";
RL Gene 81:195-202 (1989).
RN [5]
RP SEQUENCE OF 1-72 FROM N.A.
RX MEDLINE=90317835; PubMed=2164585;
RA Yoshioka Y., Fujita Y., Ohtsubo E.;
RT "Nucleotide sequence of the promoter-distal region of the tra operon
RT of plasmid R100, including traI (DNA helicase I) and traD genes.";
RL J. Mol. Biol. 214:39-53 (1990).
RN [6]
RP SEQUENCE OF 955-1756 FROM N.A.
RX MEDLINE=96347127; PubMed=8736534;
RA Perford S.S., Simon J., Frost L.S.;
RT "Regulation of the expression of the traD gene of the F sex factor of
RT Escherichia coli.";
RL Mol. Microbiol. 20:549-558 (1996).
RN [7]
RP FUNCTION: TRAI HAS BEEN IDENTIFIED AS DNA HELICASE I AND IT ALSO
CC HAS AN ADDITIONAL ACTIVITY OF SITE-SPECIFIC NICKING AT ORIT. DNA
CC HELICASE I IS A POTENT DNA-DEPENDENT ATPASE.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative initiation;
CC Comment=2 isoforms, traI (shown here) and traI*, are produced by
CC alternative initiation;
CC -1- SIMILARITY: STRONG TO TRAI OF PLASMID INCFII R100.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M54796; AAA98085.1; -;
DR EMBL; M54796; AAA98086.1; -;
DR EMBL; U01159; AAC44186.1; -;
DR EMBL; AP001918; BAA97974.1; -;
DR EMBL; M29254; AAA83930.1; ALT_INIT.
DR EMBL; X57430; CAA40677.1; -;
DR EMBL; U01159; AAC44187.1; -;
DR Ecogene; BG40119; traI.
KW Plasmid; Helicase; ATP-binding; DNA-binding; Conjugation; Hydrolase;
KW Alternative initiation; Complete proteome.
KW CHAIN 1 1756 TRAI PROTEIN, ISOFORM TRAI.
FT CHAIN 955 1756 TRAI PROTEIN, ISOFORM TRAI*.
FT INT MET 955 955 FOR ISOFORM TRAI*.
FT NP_BIND 992 999 ATP (POTENTIAL).

```

```

FT CONFLICT 69 74 MODGSN -> CRMAVT (IN RRF. 4).
SQ SEQUENCE 1756 AA; 192015 MW; AA07D61DB2BFD9FA CRC64;
Query Match 46.2%; Score 42; DB 1; Length 1756;
Best Local Similarity 66.7%; Pred. No. 83;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Oy 5 QOETVLDGNLII 16
Db 130 QSETVLTGNLVM 141
RESULT 14
TR12_ECOLI STANDARD; PRT; 1756 AA.
ID TR12_ECOLI
AC P22706;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE TraI protein (DNA helicase I) (EC 3.6.1.-).
GN TraI.
OS Escherichia coli.
OC Plasmid IncFII R100.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
NX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90317835; PubMed=2164585;
RA Yoshioka Y., Fujita Y., Ohtsubo E.;
RT "Nucleotide sequence of the promoter-distal region of the tra operon
RT of plasmid R100, including traI (DNA helicase I) and traD genes.";
RL J. Mol. Biol. 214:39-53 (1990).
RN [2]
RP FUNCTION: TRAI HAS BEEN IDENTIFIED AS DNA HELICASE I AND IT
CC ALSO HAS AN ADDITIONAL ACTIVITY OF SITE-SPECIFIC NICKING AT
CC ORIT. DNA HELICASE I IS A POTENT DNA-DEPENDENT ATPASE.
CC -1- SIMILARITY: STRONG TO TRAI OF PLASMID F.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X55815; CAA39337.1; -;
DR EMBL; S10660; BVRCAT.
KW Plasmid; Helicase; ATP-binding; DNA-binding; Conjugation;
KW Hydrolase.
FT NP_BIND 992 999 ATP (POTENTIAL).
SQ SEQUENCE 1756 AA; 191681 MW; B394B6614115F3 CRC64;
Query Match 46.2%; Score 42; DB 1; Length 1756;
Best Local Similarity 66.7%; Pred. No. 83;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Oy 5 QOETVLDGNLII 16
Db 130 QSETVLTGNLVM 141
RESULT 15
TUD_DROME STANDARD; PRT; 2515 AA.
ID TUD_DROME
AC P25823;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Maternal tudor protein.
GN TUD.
OS Drosophila melanogaster (Fruit fly).
OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

```

CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 CC NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92038995; PubMed=1936993;  
 RA Golumbecki G.S., Bardsley A., Tax F., Boswell R.E.;  
 RT "Tudor, a posterior-group gene of *Drosophila melanogaster*, encodes a  
 novel protein and an mRNA localized during mid-oogenesis.";  
 RL Genes Dev. 5:2060-2070(1991).  
 CC -1- FUNCTION: REQUIRED DURING OOGENESIS FOR THE FORMATION OF  
 CC PRIMORDIAL GERM CELLS AND FOR NORMAL ABDOMINAL SEGMENTATION.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT THE LIFE CYCLE.  
 CC MISCELLANEOUS: THE TUD MRNA ACCUMULATES WITHIN THE POSTERIOR  
 CC REGION OF THE DEVELOPING OOCYTE DURING THE EARLY TO MIDDLE STAGES  
 CC OF OOGENESIS.  
 CC -1- SIMILARITY: Contains 9 Tudor domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X62420; CAA44286.1; -.  
 DR PIR; A41519; A41519.  
 DR HSP; Q16637; 165V.  
 DR FLYBase; FBgn0003891; tud.  
 DR GO; GO:0019090; P:mitochondrial RNA, mitochondrial export; IMP.  
 DR GO; GO:0070315; P:pole plasm assembly; IMP.  
 DR InterPro; IPR001097; Maternal\_tudor.  
 DR InterPro; IPR002998; Tudor.  
 DR Pfam; PF00567; TUDOR; 10.  
 DR SMART; SM00333; TUDOR; 10.  
 DR PROSITE; PS50304; TUDOR; 9.  
 KM Developmental protein; Repeat.  
 FT DOMAIN 455 513  
 FT DOMAIN 641 696 TUDOR 1.  
 FT DOMAIN 1062 1122 TUDOR 2.  
 FT DOMAIN 1355 1414 TUDOR 3.  
 FT DOMAIN 1662 1718 TUDOR 4.  
 FT DOMAIN 1839 1898 TUDOR 5.  
 FT DOMAIN 2023 2082 TUDOR 6.  
 FT DOMAIN 2211 2269 TUDOR 7.  
 FT DOMAIN 2392 2451 TUDOR 8.  
 FT DOMAIN 2515 2573 TUDOR 9.  
 SQ SEQUENCE 2515 AA; 285236 MW; 683C100AD308BADA CRC64;  
 Query Match 46.2%; Score 42; DB 1; Length 2515;  
 Best Local Similarity 42.9%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

RP SEQUENCE FROM N.A.  
 RA Cobbe N., Heck M.M.S.;  
 RT "Phylogenetic analysis of SMC proteins."; [2]  
 RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20057165; PubMed=10591208;  
 RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,  
 RA Clamp M., Smink L.J., Alnough R., Almeida J.P., Babbage A.K.,  
 RA Baguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,  
 RA Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,  
 RA Burtil W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,  
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,  
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
 RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R., Ellington A.G.,  
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,  
 RA Gilbert J.G.R., Goward M.B., Graham D.V., Griffiths M.N.D., Hall C.,  
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,  
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,  
 RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,  
 RA Martin I.D., Mareshgh-Mohammadi M., Matthews L.H., McCann O.T.,  
 RA McElay J., McLaren S., McMurtry A.A., Milne S.A., Mortimore B.J.,  
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers J., Ross M.T.,  
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,  
 RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,  
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,  
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,  
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
 RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,  
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,  
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,  
 RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kerton S., Lai H., Lao H.I.,  
 RA Lewis J., Lewis S., Lin S.-P., Lon P., Malat J., Nguyen T., Pan H.,  
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,  
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,  
 RA Zhan M., Zhang G., Chisose S., Murray J., Miller N., Mix P.,  
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,  
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozerky P., Rolfing T.,  
 RA Scheet P., Walker C., Wamsley A., Wohlfahrt P., Pepin K., Nelson J.,  
 RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,  
 RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,  
 RA Edmund H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,  
 RA Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyraud M., Kedra D.,  
 RA Seroussi E., Franssen I., Tapia I., Bruder C.E., O'Brien K.P.,  
 RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,  
 RA Tiliun Y., Wright H.;  
 RT "The DNA sequence of human chromosome 22.";  
 RL Nature 402:489-495(1999).  
 CC -1- FUNCTION: Meiosis specific component of cohesin complex. The  
 CC cohesin complex is required for the cohesion of sister chromatids  
 CC after DNA replication. The cohesin complex apparently forms a  
 CC large proteinaceous ring within which sister chromatids can be  
 CC trapped. At anaphase, the complex is cleaved and dissociates from  
 CC chromatin, allowing sister chromatids to segregate. The meiosis-  
 CC specific cohesin complex probably replaces mitosis specific  
 CC cohesin complex when it dissociates from chromatin during prophase  
 CC I (By similarity).  
 CC -1- SUBUNIT: Forms a heterodimer with SMC3. Component of a  
 CC meiosis-specific cohesin complex, probably composed of the SMC2L2  
 CC and SMC3 heterodimer attached via their hinge domain, RAD21 (or  
 CC its meiosis-specific related protein REC8), which link them, and  
 CC SYNG3, which interacts with RAD21 or REC8 (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Nuclear protein. Associates with chromatin.  
 CC In prophase I stage of meiosis, it is found along the axial  
 CC elements of synaptonemal complexes. In late-pachytene-diplotene,  
 CC the bulk of protein dissociates from the chromosome arms probably  
 CC because of phosphorylation by PK, except at centromeres, where  
 CC cohesin complexes remain. It however remains chromatin associated  
 CC at the centromeres up to metaphase II. At anaphase II, it  
 CC dissociates from centromeres, allowing chromosomes segregation (By  
 CC similarity).

-1- DOMAIN: The flexible hinge domain, which separates the large intramolecular coiled coil regions, allows the heterotypic interaction with the corresponding domain of SMC3, forming a V-shaped heterodimer. The two heads of the heterodimer are then connected by different ends of the cleavable RAD21 or REC8 protein, forming a ring structure (by similarity).

-1- SIMILARITY: Belongs to the SMC family. SMC1 subfamily.

-1- CAUTION: Ref.2 sequence differs from that shown due to erroneous gene model prediction.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL; AJ504806; CAD3404.1; -.

DR EMBL; AL008718; -; NOT ANNOTATED CDS.

DR EMBL; AL021391; CAB41703.1; ALT\_SEQ.

DR Genew; HGNC:11112; SMC1L2.

DR InterPro; IPR003405; SMC\_C.

DR InterPro; IPR003395; SMC\_N.

DR Pfam; PF02483; SMC\_C; 1.

DR Pfam; PF02463; SMC\_N; 1.

DR Meiosis; Cell cycle; Chromosome partition; ATP-binding; Coiled coil; Nuclear protein.

KW NP BIND

FT DOMAIN 32 39 ATP (POTENTIAL).

FT DOMAIN 156 490 COILED COIL (POTENTIAL).

FT DOMAIN 491 665 FLEXIBLE HINGE.

FT DOMAIN 666 814 COILED COIL (POTENTIAL).

FT DOMAIN 851 1050 COILED COIL (POTENTIAL).

FT DOMAIN 1126 1161 ALA/ASP-RICH (DA-BOX).

FT CONFLICT 99 99 L->R (IN REF. 2).

SQ SEQUENCE 1237 AA; 144192 MW; 801DA16F447430 CRC64;

Query Match 45.6%; Score 41.5; DB 1; Length 1237;  
Best Local Similarity 56.2%; Pred. No. 68;  
Matches 9; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

Qy 3 PEOQETVLDGNLIR 17  
Db 633 PERQTVALDGTFLK 648

RESULT 17

YAO2\_SCHPO STANDARD; PRT; 150 AA.

ID\_YAO2\_SCHPO

AC Q10081;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical protein C11D3.02c in chromosome 1.

GN SPAC11D3.02C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC Schizosaccharomycetes.

OX NCBI\_TaxID=4896;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RX MEDLINE=21848401; PubMed=11859360;

RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A., Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S., Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M., Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A., Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K., James K., Jones L., Jones M., Leather S., McDonald S., McLean J., Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C., Oliver K., O'Neill S., Mungall K., Murphy L., Niblett D., Odell C.,

RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., Skelton J., Simmonds M., Squares R., Squares S., Stevens K., Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., Woodward J., Voiclaert G., Aert R., Roben J., Grymopiez B., Weltjens I., Vanstreels E., Rieger M., Schefer M., Mueller-Auer S., Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H., Borzym K., Langer I., Beck A., Leinach H., Reinhardt R., Pohl T.M., Eger P., Zimmermann W., Medler H., Wambut R., Punelle B., Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mortier S., Galibert F., Aves S.J., Xiang Z., Hunt K., Moore K., Hurst S.M., Lucas M., Roeder M., Gallardin C., Tallada V.A., Garzon A., Thode G., Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., Dominguez A., Revuelta J.U., Moreno S., Armstrong J., Forsburg S.L., Cerruti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J., Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.; "The genome sequence of Schizosaccharomyces pombe.";

RL Nature 415:871-880(2002)

CC -1- SIMILARITY: BELONGS TO THE UPF0039 (ELAA) FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL; Z68166; CAA92303.1; -.

DR PIR; T37513; T37513.

DR Genedb\_Spomb; SPAC11D3.02c; -.

DR InterPro; IPR000182; GCN5acetyltransf.

DR Pfam; PF00583; Acetyltransf; 1.

DE Hypothetical protein.

SQ SEQUENCE 150 AA; 17414 MW; A9201DD756C4CF2A CRC64;

Query Match 45.1%; Score 41; DB 1; Length 150;  
Best Local Similarity 50.0%; Pred. No. 7.6;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 3 PEOQETVLDGNLIR 16  
Db 68 PEOQETVLDGNLIR 81

RESULT 18

ARGC\_THETN STANDARD; PRT; 344 AA.

ID\_ARGC\_THETN

AC Q8R7B8;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE N-acetyl-gamma-glutamyl-phosphate reductase (EC 1.2.1.38) (AGPR) (N-acetyl-glutamate semialdehyde dehydrogenase) (NAGSA dehydrogenase).

GN ARGC OR TTE2498.

OS Thermoanaerobacter tengcongensis.

OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;

OC Thermoanaerobacteriaceae; Thermoanaerobacter.

OX NCBI\_TaxID=119072;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MB4 / JCM 11007;

RX MEDLINE=21992816; PubMed=11997336;

RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J., Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong L., Dong X., Ma Y., Ling L., Tan H., Chen R., Wang J., Yu J., Yang H.; "A complete sequence of T. tengcongensis genome.";

RL Genome Res. 12:689-700(2002).

CC -1- CATALYTIC ACTIVITY: N-acetyl-L-glutamate 5-semialdehyde + NADP(+) + phosphate = N-acetyl-5-glutamyl phosphate + NADPH.

CC -1- PATHWAY: Arginine biosynthesis, third step.

CC -1- SIMILARITY: Belongs to the NAGSA dehydrogenase family. Subfamily 1.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE013191; AAM25628.1; -.
DR HAMAP; MF_00150; -.
DR InterPro; IPR000706; AGPR_act_site.
DR InterPro; IPR000534; Semaladh_dh.
DR Pfam; PF01118; Semaladh_dh; 1.
DR Pfam; PF02774; Semaladh_dh; 1.
DR ProDom; PD003765; AGPR_act_site; 1.
DR PROSITE; PS01224; ARGC_1.
DR Arginine biosynthesis; Oxidoreductase; NADP; Complete proteome.
KM ACT_SITE 149 149 BY SIMILARITY.
SQ SEQUENCE 344 AA; 37943 MW; DCC295E5AE938CA CRC64;

Query Match 45.1%; Score 41; DB 1; Length 344;
Best Local Similarity 57.1%; Pred. No. 19;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 PEOQETVLQDNLII 16
Db 160 PLKKEKVIDGNII 173

RESULT 19
LEU3 SYNY3 STANDARD; PRT; 362 AA.
AC P73960;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 3-isopropylmalate dehydrogenase (EC 1.1.1.85) (Beta-IPM dehydrogenase)
GN (IMDH) (3-IPM-DH).
DE LEUB OR SLR1517.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=9905231.
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugitani M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- FUNCTION: Catalyzes the oxidation of 3-carboxy-2-hydroxy-4-
CC methylpentanoate (3-isopropylmalate) to 3-carboxy-4-methyl-2-
CC oxopentanoate. The product decarboxylates to 4-methyl-2-
CC oxopentanoate.
CC -1- CATALYTIC ACTIVITY: 3-carboxy-2-hydroxy-4-methylpentanoate +
CC NAD(+) = 3-carboxy-4-methyl-2-oxopentanoate + NADH.
CC -1- PATHWAY: Leucine biosynthesis; third step.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE ISOCITRATE AND ISOPROPYLMALATE
CC DEHYDROGENASES FAMILY. LEUB SUBFAMILY 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL; D90911; BAA18028.1; -.
DR PIR; S75467; S75467.
DR HSP; Q56268; 1A05.
DR HAMAP; MF_01033; -.
DR InterPro; IPR001804; Isodh.
DR InterPro; IPR004429; Leub.
DR Pfam; PF00180; Isodh; 1.
DR TIGRPFAMs; TIGR00169; leub; 1.
DR PROSITE; PS00470; IDH_IMDH; 1.
DR Oxidoreductase; Leucine biosynthesis; NAD; Complete proteome.
KM OXIDOREDUCTASE; Leucine biosynthesis; NAD; Complete proteome.
SQ SEQUENCE 362 AA; 38667 MW; F173B4112D0B8BD CRC64;

Query Match 45.1%; Score 41; DB 1; Length 362;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 KSPEQETVLQDNL 14
Db 234 RSPRQFTIVTGNL 247

RESULT 20
PBP4 HAEIN STANDARD; PRT; 479 AA.
ID P45161;
AC P45161;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Penicillin-binding protein 4 precursor (PBP-4) [includes: D-alanyl-D-
DE alanine carboxypeptidase (EC 3.4.16.4) (DD-peptidase) (DD-
DE carboxypeptidase), D-alanyl-D-alanine endopeptidase (EC 3.4.99.-) (DD-
DE endopeptidase)].
GN DABC OR H11330.
GN Haemophilus influenzae.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=9530630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -1- FUNCTION: Not involved in transpeptidation but exclusively
CC catalyzes a DD-carboxypeptidase and DD-endopeptidase reaction (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: D-alanyl-D-alanine + H(2)O = 2 D-alanine.
CC -1- PATHWAY: Peptidoglycan synthesis; final stages.
CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S13; ALSO KNOWN AS THE
CC D-ALANYL-D-ALANINE CARBOXYPEPTIDASE 3 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32812; AAC22975.1; -.
DR PIR; A64117; A64117.

```



```

DR MEROPS; S13.001; -.
DR TIGR; H11301; -.
DR InterPro; IPR000667; Peptidase_S13.
DR Pfam; PF02113; Peptidase_S13; 1.
DR PRINTS; PR00922; DADCBPTASE3.
DR TIGRPFAMS; TIGR00666; PPP4; 1.
KM Peptidoglycan synthetase; Cell division; Cell wall; Hydrolyase; Signal;
KM Antibiotic resistance; Periplasmic; Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 479
FT ACT_SITE 69 69
FT ACT_SITE 69 69
FT ACT_SITE 420 420
FT SEQUENCE 479 AA; 52685 MW; 632868C61206CB48 CRC64;

Query Match 45.1%; Score 41; DB 1; Length 479;
Best Local Similarity 77.8%; Pred. No. 28;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LDGMLIRY 18
Db 104 LDGMLIRVF 112

RESULT 21
UTGC CAEEL STANDARD; PRT; 524 AA.
ID UTGC CAEEL
AC Q22181;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Putative UDP-glucuronosyltransferase Ugt12 precursor (EC 2.4.1.17)
DE (UDPgt).
GN UGT12 OR T04H1.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=Bristol NZ;
RA Harris B.R.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP REVISIONS.
RA Dubin R.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor
CC beta-D-glucuronoside.
CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; Z78200; CAB01585.2; -.
DR PIR; T24478; T24478.
DR WormPep; T04H1.8; CB31985.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT_1.
DR PROSITE; PS00375; UDPGT_1.
KM Hypothetical protein; transferase; Glycosyltransferase; Signal;
KM Transmembrane; Glycoprotein; Multigene family.
FT SIGNAL 1 20
FT CHAIN 21 524
FT PUTATIVE UDP-GLUCURONOSYLTRANSFERASE
FT UGT12.
FT TRANSMEM 488 508
FT CARBOHYD 125 125
FT CARBOHYD 277 277
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 335 335
SQ SEQUENCE 524 AA; 59283 MW; 2CDAB5E24BDBF3CC CRC64;

Query Match 45.1%; Score 41; DB 1; Length 524;
Best Local Similarity 54.5%; Pred. No. 31;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 2 SPEQOETVLDG 12
Db 100 SPEQOETVMEG 110

RESULT 22
ID GPII RALSO STANDARD; PRT; 539 AA.
AC Q8XYN9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glucose-6-phosphate isomerase (EC 5.3.1.9) (GPI) (Phosphoglucose
DE isomerase) (PGI) (Phosphohexose isomerase) (PHI).
GN PGI OR RSC1719 OR RS02909.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Guzy J., Manganot S.,
RA Ariat M., Billault A., Brotier P., Camus J.C., Catalicio L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Signier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
CC -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate = D-fructose 6-
CC phosphate.
CC -1- PATHWAY: Involved in glycolysis and in gluconeogenesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE GPI FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; AL64066; CAB015421.1; -.
DR HAMAP; MF_00473; -.
DR InterPro; IPR01672; G6P_Isomerase.
DR Pfam; PF00342; PGI_1.
DR PRINTS; PR00662; GPIISOMERASE.
DR PROSITE; PS00765; P_GLU6OSE_ISOMERASE_1; 1.
DR PROSITE; PS00174; P_GLU6OSE_ISOMERASE_2; 1.
KM Isomerase; Gluconeogenesis; Glycolysis; Complete proteome.
FT ACT_SITE 384 384
FT ACT_SITE 505 505
FT SEQUENCE 539 AA; 58630 MW; BP3CGF1FA700BC99 CRC64;

Query Match 45.1%; Score 41; DB 1; Length 539;
Best Local Similarity 80.0%; Pred. No. 32;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 3 PEOOETVLDG 12
Db 285 PEOOETVLDG 294

```



RESULT 23  
CH60 CHLTR STANDARD; PRT; 543 AA.  
ID CH60 CHLTR  
AC 059322; P56837;  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 60 kDa chaperonin (Protein Cpn60) (GroEL protein) (Heat shock protein 60) (HSP60).  
GN GROEL OR GROEL OR MOXA OR TC0386.  
OS Chlamydia muridarum.  
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_Taxid=83560;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MOPn;  
RX MEDLINE=94215880; PubMed=7909303;  
RA Ho Y., Zhang Y.-X.;  
RT "The sequence of the groES and groEL genes from the mouse pneumonitis agent of Chlamydia trachomatis.";  
RL Gene 141:143-144 (1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MOPn;  
RA Tan M., Wong B., Engel J.N.;  
RT "The transcriptional organization and regulation of the groE and dnaK operons of Chlamydia trachomatis.";  
RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MOPn / Nig9;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Uterback T., Berry K.,  
RA Bas S., Linher K., Weidman J., Khouri H., Crahan B., Bowman C.,  
RA Dodson R., Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G.,  
RA Salzberg S.L., Bisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.";  
RL Nucleic Acids Res. 28:1397-1406 (2000).  
RN [4]  
RP FUNCTION: Prevents misfolding and promotes the refolding and proper assembly of unfolded polypeptides generated under stress conditions (By similarity).  
CC -1- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of 7 subunits (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -1- INDUCTION: By stress.  
CC -1- SIMILARITY: Belongs to the chaperonin (HSP60) family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL, L12004; AAA19871.1; -;  
DR EMBL, U52049; AAA97911.1; -;  
DR EMBL, AE002305; AAF39243.1; -;  
DR PIR; D81709; D81709.  
DR HSSP; P06139; 1GRL.  
DR TIGR; TC0386; -;  
DR HAMAP; MF\_00600; -; 1.  
DR InterPro; IPR001844; Chaperin Cpn60.  
DR InterPro; IPR002423; Cpn60/TCF-1.  
DR Pfam; PF00118; Cpn60\_TCF1; 1.  
DR PRINTS; PR00298; CHAPERONIN60.  
DR PRINTS; PR00304; TCOMPLEXCP1.  
DR PROSITE; PS00296; CHAPERONIN\_CPN60; 1.  
KM Chaperone; ATP-binding; Antigen; Heat shock; Complete proteome.  
FT INIT MET 0 0 By SIMILARITY.  
FT CONFLICT 386 386 I -> N (IN REF. 1).

SQ SEQUENCE 543 AA; 57974 MW; 41B31289CE0B271C CRC64;  
Query Match 45.1%; Score 41; DB 1; Length 543;  
Best Local Similarity 47.1%; Pred. No. 32;  
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;  
QY 2 SPEQETVLDGNLIIRY 18  
DB 206 NPTQECVLEDAVLIRY 222  
RESULT 24  
CH60 CHLTR STANDARD; PRT; 543 AA.  
ID CH60 CHLTR  
AC P17203; O84112;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE 60 kDa chaperonin (Protein Cpn60) (GroEL protein) (57 kDa chlamydial hypersensitivity antigen) (Heat shock protein 60) (HSP60).  
GN GROEL OR GROEL OR MOXA OR HYPB OR CT110.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_Taxid=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HAR-13 / Serovar A;  
RX MEDLINE=90316704; PubMed=2196231.  
RA Morrison R.P., Su H., Lyng K., Yuan Y.;  
RT "The Chlamydia trachomatis hyp operon is homologous to the groE stress operon of Escherichia coli.";  
RL Infect. Immun. 58:2701-2705 (1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Serovar L2;  
RX MEDLINE=91100013; PubMed=1987066;  
RA Cerrone M.C., Ma J.J., Stephens R.S.;  
RT "Cloning and sequence of the gene for heat shock protein 60 from Chlamydia trachomatis and immunological reactivity of the protein.";  
RL Infect. Immun. 59:79-90 (1991).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D/UM-3/Cx;  
RX MEDLINE=99000809; PubMed=9784136.  
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
RA Mitchell W.P., Olinger L., Tatunov R.L., Zhao Q., Koonin E.V.,  
RA Davis R.W.;  
RT "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.";  
RL Science 282:754-759 (1998).  
RN [4]  
RP SEQUENCE OF 1-10.  
RC STRAIN=L2/434/Bu;  
RA Bini L., Santucci A., Magi B., Marzocchi B., Sanchez-Campillo M.,  
RA Cammuccini M., Christianen G., Birklund S., Vretou E., Ratti G.,  
RA Pallini V.;  
RT Submitted (SEP-1994) to the SWISS-PROT data bank.  
RN [5]  
RP FUNCTION: Prevents misfolding and promotes the refolding and proper assembly of unfolded polypeptides generated under stress conditions (By similarity).  
CC -1- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of 7 subunits (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- INDUCTION: By stress.  
CC -1- DISEASE: THIS PROTEIN IS IMPLICATED IN THE PATHOGENESIS OF CHLAMYDIAL DISEASE. INFLAMMATION ELICITED BY THE 57 KDa ANTIGEN MAY DAMAGE TISSUE, WITH PROGRESSION TO SCARRING OF CONJUNCTIVAL AND PALLOPAIN TUBE MUCOSAE, WHICH RESPECTIVELY RESULT IN BLINDNESS AND INFERTILITY.  
CC -1- SIMILARITY: Belongs to the chaperonin (HSP60) family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M31739; AAA03204.1; -.
DR EMBL; M58027; AAA23128.1; -.
DR EMBL; AE001285; AAC67701.1; -.
DR PIR; A71555; A71555.
DR PDB; 1ROJ; 30-SEP-94.
DR PDB; 1ROK; 30-SEP-94.
DR Siena-2DPAGE; P17203; -.
DR HAMAP; MF_00600; -.
DR InterPro; IPR001844; Chaperin Cpn60.
DR InterPro; IPR002423; Cpn60/TCF-1.
DR Pfam; PF00118; Cpn60 TCPL1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCPL.
DR PROSITE; PS00296; CHAPERONIN_CPN60; 1.
KW Chaperone; ATP-binding; Antigen; Heat shock; Complete proteome;
KW 3D-structure.
FT INT MET 0 0
FT VARIANT 123 123 V -> A (IN SEROVAR L2).
FT VARIANT 130 130 I -> V (IN SEROVAR L2).
FT VARIANT 131 131 R -> K (IN SEROVAR L2 AND ISOLATE HAR-
FT 13).
FT VARIANT 188 188 I -> V (IN SEROVAR L2 AND ISOLATE HAR-
FT 13).
FT VARIANT 190 190 E -> D (IN SEROVAR L2).
FT VARIANT 216 216 D -> E (IN SEROVAR L2).
FT VARIANT 235 235 V -> I (IN ISOLATE HAR-13).
FT VARIANT 254 254 E -> V (IN SEROVAR L2).
FT VARIANT 263 263 V -> G (IN SEROVAR L2).
FT VARIANT 288 288 L -> F (IN SEROVAR L2).
FT CONFLICT 3 3 L -> F (IN REF. 4).
SQ SEQUENCE 543 AA; 58015 MW; 512161E4B6E63E CRC64;

Query Match 45.1%; Score 41; DB 1; Length 543;
Best Local Similarity 47.1%; Pred. No. 32;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SPEQETVLDGNLIIR 18
Db 206 NPETOECVLEDAVLIV 222

RESULT 25
YLIA_ECOLI STANDARD; PRT; 629 AA.
ID YLIA_ECOLI
AC P75796; Q9R3H8;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical ABC transporter ATP-binding protein ylia.
GN YLIA OR B0829.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
NC NCB1_TaxID=562;
NX NCB1_TaxID=562;
RX SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;

```

```

RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayaishi K., Horjo A.,
RA Kishimoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kishimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horikuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155 (1996).
CC -1- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM YLIA6CD. PROBABLY RESPONSIBLE FOR ENERGY COUPLING TO THE
CC TRANSPORT SYSTEM.
CC -1- SIMILARITY: Belongs to the ABC transporter family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000185; AAC73916.1; ALT_INT.
DR EMBL; D90720; BAA3517.1; ALT_INT.
DR EMBL; D90721; BAA3524.1; ALT_INT.
DR EcoGene; EG13472; YLIA.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR003439; ABC transporter.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD00006; ABC_transporter; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 2.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 2.
KW Hypothetical protein; ATP-binding; Transport; Repeat;
KW Complete proteome.
FT NP_BIND 55 62 ATP (POTENTIAL).
FT NP_BIND 363 370 ATP (POTENTIAL).
SQ SEQUENCE 629 AA; 69738 MW; 897A0D05778530CD CRC64;

Query Match 45.1%; Score 41; DB 1; Length 629;
Best Local Similarity 46.7%; Pred. No. 38;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 3 PEQETVLDGNLIIR 17
Db 307 PIEQTVVDGEPIVLR 321

RESULT 26
TETS_LISMO STANDARD; PRT; 641 AA.
ID TETS_LISMO
AC Q48791;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Tetracycline resistance protein tets (Tet(S)).
GN TETS OR TET(S).
OS Listeria monocytogenes.
OC Plasmid pIP811.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
NX NCB1_TaxID=1639;
NX NCB1_TaxID=1639;
RX SEQUENCE FROM N.A.
RC STRAIN=BM4210;
MEDLINE=93380670; PubMed=8370538;
RA Charpentier E., Gerbaud G., Courvaill P.;
RT "Characterization of a new class of tetracycline-resistance gene
RT tet(S) in Listeria monocytogenes BM4210."
RL Gene 131:27-34 (1993).
RN [2]
RP -1- FUNCTION: Abolishes the inhibitory effect of tetracyclin on
RP protein synthesis by a non-covalent modification of the ribosomes.
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.

```

```

CC      TETM/TETO SUBFAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sb-sib.ch).
CC      -----
DR      EMBL; L09756; AAA25293.1; -.
DR      PIR; JN0800; JN0800.
DR      HSSP; P13551; IDAR.
DR      InterPro; IPR000795; EF_GTPbind.
DR      InterPro; IPR000640; EFG_C.
DR      InterPro; IPR005517; EFG_IV.
DR      InterPro; IPR004161; EFTU_D2.
DR      InterPro; IPR005225; Small_GTP.
DR      Pfam; PF00679; EFG_C; 1.
DR      Pfam; PF03764; EFG_IV; 1.
DR      Pfam; PF00009; GTP_EFTU; 1.
DR      Pfam; PF03144; GTP_EFTU_D2; 1.
DR      PRINTS; PR00315; ELONGACTINCT.
DR      TIGRfams; TIGR00231; small_GTP; 1.
DR      PROSITE; PS00301; EFACOR_GTP; 1.
DR      Protein biosynthesis; Antibiotic resistance; GTP-binding; Plasmid.
FT      NP_BIND 10 17 GTP (BY SIMILARITY).
FT      NP_BIND 74 78 GTP (BY SIMILARITY).
FT      NP_BIND 128 131 GTP (BY SIMILARITY).
SQ      SEQUENCE 641 AA; 73014 MW; EC534FD38FD54FC2 CRC64;

Query March 45.1%; Score 41; DB 1; Length 641;
Best Local Similarity 55.6%; Pred. No. 39;
Matches 10; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY      3 PEOGETVLDDGN--LIIRY 18
Db      172 PEOGETVIVGNDVLLKRY 189

RESULT 27
TETS LACLA STANDARD; PRT; 646 AA.
ID TETS LACLA STANDARD; PRT; 646 AA.
AC 048712;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Tetracycline resistance protein tets (Tet(S)).
GN TETS OR TET(S).
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN NCBI_TaxID=1360;
RP SEQUENCE FROM N.A.
RC STRAIN=K214;
RA Berthelin V., Schwarz F., Cresta L., Boeglin M., Dassen G., Teuber M.;
RT "Antibiotic resistance spread in food.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Abolishes the inhibitory effect of tetracycline on
CC protein synthesis by a non-covalent modification of the ribosomes.
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC TETM/TETO SUBFAMILY.
CC -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sb-sib.ch).
CC      -----
DR      EMBL; X92946; CAA63528.1; -.

```

```

DR      HSSP; P13551; 1FNM.
DR      InterPro; IPR000795; EF_GTPbind.
DR      InterPro; IPR000640; EFG_C.
DR      InterPro; IPR005517; EFG_IV.
DR      InterPro; IPR004161; EFTU_D2.
DR      InterPro; IPR005225; Small_GTP.
DR      Pfam; PF00679; EFG_C; 1.
DR      Pfam; PF03764; EFG_IV; 1.
DR      Pfam; PF00009; GTP_EFTU; 1.
DR      Pfam; PF03144; GTP_EFTU_D2; 1.
DR      PRINTS; PR00315; ELONGACTINCT.
DR      TIGRfams; TIGR00231; small_GTP; 1.
DR      PROSITE; PS00301; EFACOR_GTP; 1.
DR      Protein biosynthesis; Antibiotic resistance; GTP-binding; Plasmid.
FT      NP_BIND 15 22 GTP (BY SIMILARITY).
FT      NP_BIND 79 83 GTP (BY SIMILARITY).
FT      NP_BIND 133 136 GTP (BY SIMILARITY).
SQ      SEQUENCE 646 AA; 73711 MW; 1F3513BA2F644B6B CRC64;

Query March 45.1%; Score 41; DB 1; Length 646;
Best Local Similarity 55.6%; Pred. No. 40;
Matches 10; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY      3 PEOGETVLDDGN--LIIRY 18
Db      177 PEOGETVIVGNDVLLKRY 194

RESULT 28
VG50 BPML5 STANDARD; PRT; 682 AA.
ID VG50 BPML5 STANDARD; PRT; 682 AA.
AC 005262;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Putative adenoviral-cobalamin-dependent ribonucleotide reductase
DE (EC 1.17.4.2) (GP50).
GN 50.
OS Mycobacteriophage L5.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC L5-like viruses.
OX NCBI_TaxID=31757;
RN NCBI_TaxID=31757;
RP SEQUENCE FROM N.A.
RX MEDLINE=93211282; PubMed=8459766;
RA Hatfull G.F., Sarkis G.J.;
RT "DNA sequence, structure and gene expression of mycobacteriophage L5:
RT a phage system for mycobacterial genetics.";
RL Mol. Microbiol. 7:395-405(1993).
CC -1- CATALYTIC ACTIVITY: 2'-deoxyribonucleoside triphosphate + oxidized
CC thiorodoxin + H(2)O = ribonucleoside triphosphate + reduced
CC thiorodoxin.
CC -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sb-sib.ch).
CC      -----
DR      EMBL; Z18946; CAA79426.1; -.
DR      PIR; S30995; S30995.
DR      InterPro; IPR000788; Ribonucleo_red.
DR      Pfam; PF02867; ribonuc_red_1gc; 1.
DR      Oxidoreductase; DNA replication.
SQ      SEQUENCE 682 AA; 76331 MW; 8D2A71B873BC04A8 CRC64;

Query March 45.1%; Score 41; DB 1; Length 682;
Best Local Similarity 57.1%; Pred. No. 42;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      3 PEOGETVLDDGNLIIRI 16

```

RESULT	29
ID	ITTH2_MOUSE
AC	061703;
DT	15-JUL-1998 (Rel. 36, Created)
DT	15-JUL-1998 (Rel. 36, Last sequence update)
DT	28-FEB-2003 (Rel. 44, Last annotation update)
DE	Inter-alpha-trypsin inhibitor heavy chain H2 precursor (ITI heavy chain H2) (inter-alpha-inhibitor heavy chain 2).
GN	ITTH2.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=C57BL/6N; TISSUE=Liver;
RX	MEDLINE=95194326; PubMed=7534067;
RA	Chan P., Risler J.-L., Raugen G., Sailer J.-P.;
RT	"The three heavy-chain precursors for the inter-alpha-inhibitor family in mouse: new members of the multicopper oxidase protein group with differential transcription in liver and brain."
RT	Biochem. J. 306:505-512(1995).
RL	-1- FUNCTION: MAY ACT AS A CARRIER OF HYALURONAN IN SERUM OR AS A BINDING PROTEIN BETWEEN HYALURONAN AND OTHER MATRIX PROTEIN, INCLUDING THOSE ON CELL SURFACES IN TISSUES TO REGULATE THE LOCALIZATION, SYNTHESIS AND DEGRADATION OF HYALURONAN WHICH ARE ESSENTIAL TO CELLS UNDERGOING BIOLOGICAL PROCESSES (BY SIMILARITY).
CC	-1- SUBUNIT: I-ALPHA-I PLASMA PROTEASE INHIBITORS ARE ASSEMBLED FROM ONE OR TWO HEAVY CHAINS (H1, H2 OR H3) AND ONE LIGHT CHAIN, BIKUNIN. INTER-ALPHA-INHIBITOR (I-ALPHA-I) IS COMPOSED OF H1, H2 AND BIKUNIN. INTER-ALPHA-LIKE INHIBITOR (I-ALPHA-LI) OF H2 AND BIKUNIN, AND PRE-ALPHA-INHIBITOR (P-ALPHA-I) OF H3 AND BIKUNIN.
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN BOTH LIVER AND BRAIN.
CC	-1- PFM: HEAVY CHAINS ARE INTERLINKED WITH BIKUNIN VIA A CHONDROITIN 4-SULFATE BRIDGE TO THE THEIR C-TERMINAL ASPARTATE (BY SIMILARITY).
CC	-1- SIMILARITY: BELONGS TO THE ITTH FAMILY.
CC	-1- SIMILARITY: Contains 1 WMPA domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as their content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; X70392; CAA49842.1; -.
DR	PIR; S54354; S54354.
DR	MGI; MGI:96619; Itih2.
DR	InterPro; IPR006587; VIP.
DR	InterPro; IPR002035; WVF_A.
DR	Pfam; PF00092; wvf_1.
DR	SMART; SM00609; VIT; 1.
DR	SMART; SM00327; VMA; 1.
DR	PROSITE; PS50234; WVFA; 1.
KW	Serine protease inhibitor; Repeat; Signal; Multigene family; Glycoprotein.
FT	1
FT	18
FT	PROTEIN.
FT	PROPEP
FT	19
FT	54
FT	BY SIMILARITY.
FT	INTER-ALPHA-TRYPSIN INHIBITOR HEAVY CHAIN
FT	H2.
FT	BY SIMILARITY.
FT	WVFA.
FT	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	N-LINKED (GLCNAc. . .) (POTENTIAL).
FT	CARBOHYD
FT	445
FT	445
FT	445

FT	BINDING	702	702		CHONDROITIN-4-SULFATE, CROSS-LINK SITE (BY SIMILARITY)
SQ	SEQUENCE	946 AA;	105927 MW;	40D6B716433ED5DC CRC64;	
	Query Match		45.1%;	Score 41; DB 1; Length 946;	
	Best Local Similarity		38.9%;	Pred. No. 61;	
	Matches	7; Conservative	4; Mismatches	7; Indels	0; Gaps
OY		1 KSPROETVLDGNIIRY 18			
Dd		260 KCPCNCTETAANGVELVMY 277			
<hr/>					
RESULT 30					
INR1_HUMAN	STANDARD;	PRT;	557 AA.		
ID INR1_HUMAN	PI181;				
DT 01-AUG-1990	(Rel. 15, Created)				
DT 01-AUG-1990	(Rel. 15, Last sequence update)				
DT 16-OCT-2001	(Rel. 40, Last annotation update)				
DE Interferon-alpha/beta receptor alpha chain precursor (IFN-alpha-REC).					
GN IFNARI OR IFNAR.					
OS Homo sapiens (Human).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
OX NCBI_TaxID=9606;					
RN [1]					
RP SEQUENCE FROM N.A.					
RX MEDLINE=90124632; PubMed=2153461;					
RA Uze G., Lutfalla G., Gresser I.;					
RT "Genetic transfer of a functional human interferon alpha receptor					
into mouse cells: cloning and expression of its cDNA.";					
RL Cell 60:225-234(1990).					
RN [2]					
RP SEQUENCE FROM N.A.					
RX MEDLINE=92129376; PubMed=1370833;					
RA Lutfalla G., Gardiner K., Proudhon D., Vielh E., Uze G.;					
RT "The structure of the human interferon alpha/beta receptor gene."					
RL J. Biol. Chem. 267:2802-2809(1992).					
[3]					
RP PHOSPHORYLATION BY TYK2.					
RX Medline=96059042; PubMed=7526154;					
RA Colamonici O., Yan H., Domanski P., Handa R., Smalley D.,					
Mullersman J., Witte W., Krishnan K., Kroleski J.;					
RT "Direct binding to and tyrosine phosphorylation of the alpha subunit					
of the type I interferon receptor by p135tyk2 tyrosine kinase.";					
RL Mol. Cell. Biol. 14:8133-8142(1994).					
-I FUNCTION: RECEPTOR FOR INTERFERONS ALPHA AND BETA. BINDING TO TYPE					
I IFNS TRIGGERS TYROSINE PHOSPHORYLATION OF A NUMBER OF PROTEINS					
INCLUDING JMKS, TYK2, STAT PROTEINS AND IFN-R ALPHA-AND BETA-					
SUBUNTS THEMSELVES.					
-I SUBCELLULAR LOCATION: Type I membrane protein.					
-I TISSUE SPECIFICITY: IFN RECEPTORS ARE PRESENT IN ALL TISSUES AND					
EVEN ON THE SURFACE OF MOST IFN-RESISTANT CELLS. TYROSINE KINASE.					
-I PTM: PHOSPHORYLATED ON TYROSINE RESIDUES BY TYK2.					
-I SIMILARITY: Contains 2 fibronectin type III domains.					
-I SIMILARITY: BELONGS TO THE CLASS II CYTOKINE FAMILY OF RECEPTORS.					
-----					
This SWISS-PROT entry is copyright. It is produced through a collaboration					
between the Swiss Institute of Bioinformatics and the EMBL outstation -					
the European Bioinformatics institute. There are no restrictions on its					
use by non-profit institutions as long as its content is in no way					
modified and this statement is not removed. Usage by and for commercial					
entities requires a license agreement (See http://www.ebi.ac.uk/announce/					
or send an email to license@ebi.ac.uk).					
-----					
CC EMBL; J03171; AAA52730.1; -					
DR EMBL; X60459; CAA42992.1; -					
DR PIR; A32694; A32694.					
DR GeneW; HGNC:5432; IFNARI.					
DR MIM; 107450; -					
DR GO; GO:0005887; C:integral to plasma membrane; TAS.					
DR GO; GO:0004905; F:interferon-alpha/beta receptor activity; TAS.					

DR GO: GO:0007166; P:cell surface receptor linked signal transdu. . .; TAS.  
 DR GO: GO:0007259; P:JAK-STAT cascade; TAS.  
 DR GO: GO:0009615; P:response to viruses; TAS.  
 DR InterPro: IPR000283; Cytok receptor\_2.  
 DR InterPro: IPR003961; FN\_III.  
 DR SMART: SM00060; FN3; 3.  
 KW Receptor; Transmembrane; Glycoprotein; Signal; Polymorphism;  
 KM Phosphorylation.  
 FT SIGNAL 1 27  
 FT CHAIN 28 557  
 FT DOMAIN 28 436  
 FT TRANSMEM 437 457  
 FT DOMAIN 458 557  
 FT DISULFID 79 87  
 FT DISULFID 199 220  
 FT MOD\_RES 466 466  
 FT MOD\_RES 481 481  
 FT CARBOHYD 50 50  
 FT CARBOHYD 58 58  
 FT CARBOHYD 81 81  
 FT CARBOHYD 88 88  
 FT CARBOHYD 110 110  
 FT CARBOHYD 172 172  
 FT CARBOHYD 254 254  
 FT CARBOHYD 313 313  
 FT CARBOHYD 314 314  
 FT CARBOHYD 376 376  
 FT CARBOHYD 416 416  
 FT CARBOHYD 433 433  
 FT VARIANT 168 168  
 FT CONFLICT 17 17  
 FT SEQUENCE 557 AA; 63525 MM; 0F6744C6A1ADB37 CRC64;  
 Query Match 44.5%; Score 40.5; DB 1; Length 557;  
 Best Local Similarity 42.1%; Pred. No. 41;  
 Matches 8; Conservative 6; Mismatches 4; Indels 1; Gaps 1;  
 QY 1 KSPBQOET-VLDGULIRY 18  
 DB 31 KSPBQKVEVDIIDNFIIRW 49  
 RESULT 31  
 V50K\_BYDV STANDARD; PRT; 450 AA.  
 ID V50K\_BYDV  
 AC P09516;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 50 kDa protein (ORF 4).  
 OS Barley yellow dwarf virus (isolate PAV) (BYDV).  
 OS Barley yellow dwarf virus (isolate PAV) (BYDV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;  
 OC Luteovirus.  
 NCBI\_TaxID=12040;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88289355; PubMed=3399386;  
 RA Miller W.A., Waterhouse P.M., Gerlach W.L.;  
 RT "Sequence and organization of barley yellow dwarf virus genomic RNA.";  
 RL Nucleic Acids Res. 16:6097-6111(1988)  
 CC -1- SIMILARITY: TO ORF 6 OF BEET WESTERN YELLOW VIRUS, ORF6 OF  
 CC POTATO LEAFROLL VIRUS AND OF 54K PROTEIN OF BEET NECROSIS  
 CC YELLOW VEIN VIRUS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL, outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL: X07653; CAA30495.2; -  
 DR InterPro: IPR002929; PLRV ORFS.  
 DR Pfam: PF01690; PLRV ORFS\_1.  
 SQ SEQUENCE 450 AA; 49765 MM; 3960B2663D45331D CRC64;  
 Query Match 44.0%; Score 40; DB 1; Length 450;  
 Best Local Similarity 54.5%; Pred. No. 39;  
 Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KSPBQOETVLD 11  
 DB 283 KTBQOETVLD 293  
 RESULT 32  
 C4DE\_DROME STANDARD; PRT; 507 AA.  
 ID C4DE\_DROME  
 AC 046051;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Probable cytochrome P450 4d14 (BC 1.14.-.-) (CYP11D14).  
 GN CYP4D14 OR BG:152A3.2 OR CG3540.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Drosophila melanogaster; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkelley.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams W.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brodeur P., Brothier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo E., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Schaefer F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Sult T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.

```

RC STRAIN=Oregon-R; PubMed=10731137;
RX MEDLINE=20196011;
RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demallert J., Cadieu E.,
RA Dreano S., Gloux S., Lelaure V., Mottier S., Gallibert F., Borkova D.,
RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
RA Papagiamakis G., Spanos L., Cox S., Madueno E., de Pablo B.,
RA Modolell J., Peter A., Schottler P., Werner M., Mourikioti F.,
RA Benet N., Dowe G., Schaefer U., Jaackle H., Bucheton A.,
RA Callister D.M., Campbell L.A., Darlambou A., Henderson N.S.,
RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,
RA Glover D.M.,
RT "From sequence to chromosome: the tip of the X chromosome of D.
RT melanogaster.";
RT Science 287:2220-2222(2000).
CC -1- FUNCTION: MAY BE INVOLVED IN THE METABOLISM OF INSECT HORMONES AND
CC IN THE BREAKDOWN OF SYNTHETIC INSECTICIDES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +
CC oxidized flavoprotein + H(2)O.
CC -1- SUBCELLULAR LOCATION: Membrane-bound; Endoplasmic reticulum
CC (Potential).
CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE003423; AAF45740.1; -.
DR EMBL; AL009194; CAA15696.1; -.
DR HSSP; P14779; 1JBP2.
DR FlyBase; FBgn0023541; Cyp4d14.
DR InterPro; IPR001128; Cytochrome_P450.
DR Pfam; PF00067; P450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;
KW Endoplasmic reticulum; Hypothetical protein;
FT METTL 454 454 IKON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 507 AA; 57480 MW; 01D33A556CA969A4 CRC64;

Query Match 44.0%; Score 40; DB 1; Length 507;
Best Local Similarity 80.0%; Pred. No. 45;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 6 QETVLDGNLI 15
DB 391 QDTVLDGKLI 400

RESULT 33
ID CH61_CHL CV STANDARD; PRT; 544 AA.
AC P15559;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 60 kDa chaperonin 1 (protein Cpn60 1) (groEL protein 1) (57 kDa
DE chlamydial hypersensitivity antigen).
GN GROEL1 OR GROEL1 OR GROEL-1 OR GROL OR GROEL OR MOPA OR HYB OR
GN CCA00643.
OS Chlamydomonas caviae.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomonada.
OX NCBI_TaxID=83557;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GPIC;
RX MEDLINE=90010784; PubMed=2571668;
RX Morrison R.P., Belland R.J., Lyons K., Caldwell H.D.;
RT "Chlamydial disease pathogenesis. The 57-kD chlamydial

```

```

RT hyperensitivity antigen is a stress response protein.";
RL J. Exp. Med. 170:1271-1283(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GPIC;
RX MEDLINE=22569155; PubMed=12682364;
RA Read T.D., Myers G.S.A., Brunham R.C., Nelson W.C., Paulsen I.T.,
RA Heidelberg J., Holtzapple E., Knouri H., Fedorova N.B., Catty H.A.,
RA Umeyam L.A., Halt D.H., Peterson J., Beaman W.J., White O.,
RA Salzberg S.L., Hsia R.-C., McCarty G., Rank R.G., Bavoil P.M.,
RA Fraser C.M.;
RT "Genome sequence of Chlamydomonas caviae (Chlamydia psittaci GPIC):
RT examining the role of niche-specific genes in the evolution of the
RT Chlamydiaceae.";
RL Nucleic Acids Res. 31:2134-2147(2003).
CC -1- FUNCTION: Prevents misfolding and promotes the refolding and
CC proper assembly of unfolded polypeptides generated under stress
CC conditions (By similarity).
CC -1- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of
CC 7 subunits (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- INDUCTION: By stress.
CC -1- DISEASE: This protein is implicated in the pathogenesis of
CC chlamydial disease. Inflammation elicited by the 57 kDa antigen
CC may damage tissue, with progression to scarring of conjunctival
CC and fallopian tube mucosae, which respectively result in
CC blindness and infertility.
CC -1- SIMILARITY: Belongs to the chaperonin (HSP60) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X51404; CAA35766.1; -.
DR EMBL; AE016966; AAF05385.1; -.
DR PIR; J0117; J0117.
DR HSSP; P06139; 1GRL.
DR TIGR; CCA00643; -.
DR HAMAP; MF_00600; -, 1.
DR InterPro; IPR001844; Chaperonin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCF-1.
DR Pfam; PF00118; Cpn60 TCPL; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; COMPLEXITCP1.
DR PROSITE; PS00296; CHAPERONINS_CPN60; 1.
DR Chaperone; ATP-binding; Antigen.
KW Chaperone; ATP-binding; Antigen.
SQ SEQUENCE 544 AA; 58154 MW; 998A5A98C914A09A CRC64;

Query Match 44.0%; Score 40; DB 1; Length 544;
Best Local Similarity 47.1%; Pred. No. 48;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 2 SPEQETVLDGNLIRY 18
DB 207 NPETQCVLBERALVILY 223

RESULT 34
ID CU23_SCHPO STANDARD; PRT; 565 AA.
AC O94556;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 20S cyclosome/Apc complex protein cut23 (Anaphase promoting factor
DE component 8).
GN CU23 OR APC8 OR SPAC6F12.14.
OS Schizosaccharomyces pombe (Fission yeast).
OC Schizosaccharomycetes; Ascomycota; Schizosaccharomycetes;
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

```

CC Schizosaccharomycetales; Schizosaccharomycetales;  
 CC Schizosaccharomycetes.  
 CC NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND FUNCTION.  
 RC STRAIN=972;  
 RX MEDLINE=99457297; PubMed=10526233;  
 RA Yamashita Y.M., Nakaseko Y., Kumada K., Nakagawa T., Yanagida M.;  
 RT "Fission yeast APC/cyclosome subunits, Cut20/apc4 and Cut23/apc8, in  
 RT regulating metaphase-anaphase progression and cellular stress  
 RT responses";  
 RT Genes Cells 4:445-463(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brookes K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Widdowson I., Widdowson E., Aert R., Robben J., Grymoultz B.,  
 RA Widdowson J., Widdowson E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moesl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,  
 RA Beer P., Zimmermann W., Medler H., Wambut R., Purnelle B.,  
 RA Coffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hunt S.M.,  
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,  
 RA Carnuti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Uesery D., Barrett B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 CC -1- FUNCTION: Has a role in promoting metaphase to anaphase transition  
 CC via the ubiquitination of specific mitotic substrates.  
 CC -1- SIMILARITY: Contains 8 TPR repeats.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: Z98533; CAB1101.2; -  
 DR GeneDB: Spombe; SPAC6P12.14; -  
 DR InterPro: IPR007192; AFC8.  
 DR InterPro: IPR001440; TPR.  
 DR Pfam: PF04049; APC8; 1.  
 DR Pfam: PF00515; TPR; 6.  
 DR Cell cycle; Microtubule; Repeat; TPR repeat.  
 FT REPEAT 87 120  
 FT REPEAT 173 206 TPR 1.  
 FT REPEAT 270 303 TPR 2.  
 FT REPEAT 338 371 TPR 3.  
 FT REPEAT 373 405 TPR 4.  
 FT REPEAT 407 439 TPR 5.  
 FT REPEAT 440 473 TPR 6.  
 FT REPEAT 475 507 TPR 7.  
 FT REPEAT 507 539 TPR 8.  
 FT SEQUENCE 565 AA; 65854 MW; 02B4C9CC07D21B80 CRC64;  
 Query Match 44.0%; Score 40; DB 1; Length 565;  
 Best Local Similarity 50.0%; Pred. No. 50;

Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 Oy 1 KSPPOETVLQDNLII 16  
 Db 133 KSEBNETLNTNLTL 148  
 RESULT 35  
 ID PIGR RAT STANDARD; PRT; 769 AA.  
 AC P15083;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Polymeric-immunoglobulin receptor precursor (poly-Ig receptor) (PIGR)  
 DE [Contains: Secretory component].  
 GN PIGR.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=89378226; PubMed=2776882;  
 RA Banting G., Brake B., Braghetta P., Luzio J.P., Stanley K.K.;  
 RT "Intracellular targeting signals of polymeric immunoglobulin  
 RT receptors are highly conserved between species";  
 RL FEBS Lett. 254:177-183(1989).  
 CC -1- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE  
 CC TRANSAPICAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN  
 CC TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.  
 CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE  
 CC EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE  
 CC TRANSMEMBRANE SEGMENT.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.  
 CC -1- SIMILARITY: Contains 5 immunoglobulin-like V-type domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X15741; CA33758.1; -  
 DR PIR: S05407; ORRTGS.  
 DR InterPro: IPR007110; Ig-like.  
 DR InterPro: IPR003599; Ig.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR Pfam: PF00047; Ig; 5.  
 DR SMART: SM00409; Ig; 5.  
 DR PROSITE: PS50835; IG LIKE; 2.  
 DR Immunoglobulin domain; Repeat; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 769  
 FT CHAIN 19 611  
 FT CHAIN 19 643  
 FT TRANSMEM 644 666  
 FT TRANSMEM 667 769  
 FT DOMAIN 21 126  
 FT DOMAIN 135 237  
 FT DOMAIN 240 341  
 FT DOMAIN 353 457  
 FT DOMAIN 463 563  
 FT DISULFID 40 110  
 FT DISULFID 152 220  
 FT DISULFID 257 324  
 FT DISULFID 370 440  
 FT DISULFID 484 546  
 FT CARBOHYD 90 135  
 FT CARBOHYD 135 135  
 POLYMERIC-IMMUNOGLOBULIN RECEPTOR.  
 SECRETORY COMPONENT.  
 EXTRACELLULAR (POTENTIAL).  
 POTENTIAL.  
 CYTOPLASMIC (POTENTIAL).  
 IG-LIKE V-TYPE 1.  
 IG-LIKE V-TYPE 2.  
 IG-LIKE V-TYPE 3.  
 IG-LIKE V-TYPE 4.  
 IG-LIKE V-TYPE 5.  
 POTENTIAL.  
 POTENTIAL.  
 POTENTIAL.  
 POTENTIAL.  
 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 N-LINKED (GLCNAC. . .) (POTENTIAL).



```

FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 769 AA; 84798 MW; 5F849303400255A7 CRC64;
Query Match 44.0%; Score 40; DB 1; Length 769;
Best Local Similarity 63.6%; Pred. No. 71;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 2 SPROGTVLADG 12
Db 706 TPDQGETVLEG 716
RESULT 36
CAR9_HUMAN STANDARD; PRT; 536 AA.
ID CAR9_HUMAN 09H854;
AC 09H854; 09H854;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Caspase recruitment domain protein 9 (hCARD9).
GN CARD9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RX MEDLINE=20576268; PubMed=11053425;
RA Bettin J., Guo Y., Wang L., Srinivasula S.M., Jacobson M.D.,
RA Poyet J.-L., Merriam S., Du M.-Q., Dyer M.J.S., Robison K.E.,
RA Dilefano P.S., Alnemri E.S.;
RT "CARD9 is a novel caspase recruitment domain-containing protein that
RT interacts with Bcl10/CLAP and activates NF-kappa B.";
RL J. Biol. Chem. 275:41082-41086(2000).
[2]
SEQUENCE FROM N.A.
RC TISSUE=Retinoblastoma;
RA Iisogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Magatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,
RA Niimoriya K., Iwayanagi T.;
RT "NEO human CDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
[3]
SEQUENCE OF 1-492 FROM N.A.
RC TISSUE=Muscle;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L.H., Scheiner C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Dichtchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stepien M., Soares M.B., Bonaldi M.F., Caavanti T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carinci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.T.,
RA Bosak S.A., McKen P.J., McKernan K.V., Malek J.A., Gunaratne P.H.,
RA Richards S., Wozny K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Paley J., Helton E., Kettman M., Madan A., Rodriques S., Sanchez A.,
RA Whiting R.W., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield A.S., Krzyzinski M.I., Skalska U., Smalios D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: Activates NF-kappaB via Bcl10.
CC -1- SUBUNIT: Self-associates. CARD9 and Bcl10 bind to each other by

```

```

CC CARD-CARD interaction.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: Highly expressed in spleen. Also detected in
CC liver, placenta, lung, peripheral blood leukocytes and in brain.
CC -1- SIMILARITY: Contains 1 CARD domain.
CC -1- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 360.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF311287; AAC82790.1; -.
DR EMBL; AK024001; BAB14766.1; ALT_FRAME.
DR EMBL; BC008877; AAH08877.1; -.
DR Genew; HGNC:16391; CARD9.
DR MIM; 607212; -.
DR InterPro; IPR001315; CARD.
DR PROSITE; PS50209; CARD; 1.
KW Coiled coil.
FT DOMAIN 6 98 CARD.
FT 117 277 COILED COIL (POTENTIAL).
FT DOMAIN 332 419 COILED COIL (POTENTIAL).
FT CONFLICT 12 12 N -> S (IN REF. 3).
FT CONFLICT 482 492 LSGEPEPEK -> PAGLGIGAVC (IN REF. 3).
SQ SEQUENCE 536 AA; 62267 MW; 6EB1835315B83D55 CRC64;
Query Match 43.4%; Score 39.5; DB 1; Length 536;
Best Local Similarity 52.9%; Pred. No. 58;
Matches 9; Conservative 4; Mismatches 3; Indels 1; Gaps 1;
QY 2 SPROGTVL-DGNLIR 17
Db 41 NPDDGTVLSDPNLIVIR 57
RESULT 37
CAR9_RAT STANDARD; PRT; 536 AA.
ID CAR9_RAT 09BPT0;
AC 09BPT0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Caspase recruitment domain protein 9 (rCARD9).
GN CARD9.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RX MEDLINE=20576268; PubMed=11053425;
RA Bettin J., Guo Y., Wang L., Srinivasula S.M., Jacobson M.D.,
RA Poyet J.-L., Merriam S., Du M.-Q., Dyer M.J.S., Robison K.E.,
RA Dilefano P.S., Alnemri E.S.;
RT "CARD9 is a novel caspase recruitment domain-containing protein that
RT interacts with Bcl10/CLAP and activates NF-kappa B.";
RL J. Biol. Chem. 275:41082-41086(2000).
[2]
SEQUENCE OF 1-492 FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=20576268; PubMed=11053425;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L.H., Scheiner C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Dichtchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stepien M., Soares M.B., Bonaldi M.F., Caavanti T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carinci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.T.,
RA Bosak S.A., McKen P.J., McKernan K.V., Malek J.A., Gunaratne P.H.,
RA Richards S., Wozny K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Paley J., Helton E., Kettman M., Madan A., Rodriques S., Sanchez A.,
RA Whiting R.W., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield A.S., Krzyzinski M.I., Skalska U., Smalios D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- FUNCTION: Activates NF-kappaB via Bcl10 (By similarity).
CC -1- SUBUNIT: Self-associates. CARD9 and Bcl10 bind to each other by
CC CARD-CARD interaction (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Contains 1 CARD domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```



```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ldb-sib.ch/announce/
CC or send an email to license@ldb-sib.ch).
CC -----
DR EMBL; AF311288; AAC28791.1; -.
DR InterPro; IPR001315; CARD.
DR PROSITE; PS50209; CARD; 1.
CC Coiled coil.
FT DOMAIN 6 98 CARD.
FT DOMAIN 117 277 COILED COIL (POTENTIAL) .
FT DOMAIN 303 420 COILED COIL (POTENTIAL) .
SQ SEQUENCE 536 AA; 62631 MW; 6F33089C87E5BAC9 CRC64;

Query Match 43.4%; Score 39.5; DB 1; Length 536;
Best Local Similarity 52.9%; Pred. No. 58;
Matches 9; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 2 SPEQGETVL-NGNLIR 17
Db 41 NPDDDEQVLSDFNLVIR 57

RESULT 38
Y082 CAEEL
ID _Y082 CAEEL STANDARD; PRT; 128 AA.
AC Q09237;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C18H9.2 in chromosome II.
GN C18H9.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditie.
OX NCBI_TaxID=62239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Favello A.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ldb-sib.ch/announce/
CC or send an email to license@ldb-sib.ch).
CC -----
CC EMBL; U23147; AAC46687.1; -.
DR PIR; B88188; B88188.
DR WormPep; C18H9.2; CE01802.
KW Hypothetical protein.
SQ SEQUENCE 128 AA; 13666 MW; C7A64D960B95E2EC CRC64;

Query Match 42.9%; Score 39; DB 1; Length 128;
Best Local Similarity 46.7%; Pred. No. 14;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 KSPEQGETVLNGNL 15
Db 85 KKKKTQKQVLDGNL 99

RESULT 39
PSAF SYNY3
ID _PSAF SYNY3 STANDARD; PRT; 165 AA.
AC P29256;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Photosystem I reaction centre subunit III precursor (PSI-F).
OS PSAF OR S1L0819.

```

```

OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxId=1148;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92041839; PubMed=1939076;
RX Chitnis P.R., Purvis D., Nelson N.;
RT "Molecular cloning and targeted mutagenesis of the gene psaf encoding
RT subunit III of photosystem I from the cyanobacterium Synchocystis
RT sp. PCC 6803.";
RL J. Biol. Chem. 266:20146-20151(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=94448808; PubMed=8106355;
RX Xu Q., Yu L., Chitnis V.P., Chitnis P.R.;
RT "Function and organization of photosystem I in a cyanobacterial
RT mutant strain that lacks Psaf and Psaf subunits.";
RL J. Biol. Chem. 269:3205-3211(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA MEDLINE=97061201; PubMed=8905231;
RX Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Hisouji M., Matsuo A., Sugihara M., Sasamoto S., Kimura T.,
RA Moench T., Hirosewa M., Muraki C., Wada T., Watanabe A.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res 3:109-136(1996).
CC -I- FUNCTION: PROBABLY PARTICIPATES IN EFFICIENCY OF ELECTRON TRANSFER
CC FROM PLASTOCYANIN TO P700 (OR CYTOCHROME C553 IN ALGAE AND
CC CYANOACETIC ACID). THIS PLASTOCYANIN-DOCKING PROTEIN CONTRIBUTES
CC TO THE SPECIFIC ASSOCIATION OF PLASTOCYANIN TO PS I.
CC -I- SIMILARITY: BELONGS TO THE PSAF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isdb.ch/announce/
CC or send an email to license@isdb-sib.ch).
CC -----
DR EMBL; M74801; AAA27292.1; -;
DR EMBL; L20938; AAA27294.1; -;
DR EMBL; D90911; BAA18108.1; -;
DR PIR; A41072; A41072.
DR InterPro; IPR003666; PSI_PSAF.
DR Pfam; PF02507; PSI_Psaf_1.
KW Photosynthesis; Photosystem I; Signal; Complete proteome.
FT SIGNAL 1 23 BY SIMILARITY.
FT CHAIN 24 165 PHOTOSYSTEM I REACTION CENTRE SUBUNIT
FT III.
SQ SEQUENCE 165 AA; 18249 MM; 9DD84F6B87E72151 CRC64;
Query Match 42.9%; Score 39; DB 1; Length 165;
Matches 8; Similarity 47.1%; Pred. No. 19;
Match 5; Indels 0; Gaps 0;
QY 1 KSPQOETVLDGNLIIR 17
DB 116 KNPEQEVIVNVPLAIK 132
RESULT 40
TRYL ANOGA STANDARD; PRF; 274 AA.
AC P35035;
DT 01-FEB-1994 (Rel. 28, Last Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Trypsin 1 precursor (EC 3.4.21.4).
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 10, 2003, 17:19:24 ; Search time 95 Seconds  
(without alignments)  
48.894 Million cell updates/sec

Title: US-09-991-795-1  
Perfect score: 91  
Sequence: 1 KSPQOQETVLDGMLIRY 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues  
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: SP archaea:\*
  - 2: SP bacteria:\*
  - 3: SP fungi:\*
  - 4: SP human:\*
  - 5: SP invertebrate:\*
  - 6: SP mammal:\*
  - 7: SP mhc:\*
  - 8: SP organelle:\*
  - 9: SP phage:\*
  - 10: SP plant:\*
  - 11: SP rodent:\*
  - 12: SP virus:\*
  - 13: SP vertebrate:\*
  - 14: SP unclassified:\*
  - 15: SP virus:\*
  - 16: SP bacteriap:\*
  - 17: SP archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	57	62.6	902	11	Q9DBK8
2	57	62.6	932	11	Q35802
3	57	62.6	941	11	Q91W60
4	57	62.6	941	11	Q8C7K5
5	57	62.6	942	11	Q54882
6	57	62.6	942	11	Q8C7G9
7	57	62.6	942	11	Q8C7G9
8	47.5	52.2	604	4	Q96CJ3
9	46	50.5	179	16	Q8E2L5
10	46	50.5	179	16	Q8E2L5
11	45	49.5	598	17	Q8DWR6
12	45	49.5	319	10	Q9AJZ1
13	45	49.5	452	10	Q9ATJ5
14	45	49.5	452	10	Q8G265
15	45	49.5	1113	11	Q8BVS2
16	44	48.4	409	10	Q39363

17	44	48.4	431	10	Q23849	023849 brassica ca
18	44	48.4	493	5	Q8MNC9	Q8mnc9 dictyosteli
19	44	48.4	856	10	Q39277	Q39277 brassica ca
20	43	47.3	401	2	Q8VR20	Q8vr20 chlamydomo
21	43	47.3	628	16	Q8NWF5	Q8nwf5 staphylococ
22	43	47.3	736	10	Q94GY4	Q94gy4 oryza sativ
23	43	47.3	749	16	Q8EL73	Q8el73 oceanobacil
24	43	47.3	858	16	Q8E2R3	Q8e2r3 streptococc
25	43	47.3	858	16	Q8DWM1	Q8dwm1 streptococc
26	42	46.2	312	10	Q9ATJ8	Q9atj8 arabidopsis
27	42	46.2	360	2	Q9EJ05	Q9ej05 salmoneilla
28	42	46.2	428	10	Q23853	Q23853 brassica ca
29	42	46.2	497	2	Q9S6B0	Q9s6b0 chlamydia p
30	42	46.2	858	10	Q01963	Q01963 brassica na
31	42	46.2	1752	16	Q93GL4	Q93gl4 salmonella
32	42	46.2	1756	2	Q9WTB0	Q9wtb0 escherichia
33	42	46.2	2515	5	Q9W2U8	Q9w2u8 drosophila
34	41.5	45.6	346	16	Q8CQA1	Q8cqa1 staphylococ
35	41.5	45.6	806	4	Q9Y3G5	Q9y3g5 homo sapien
36	41.5	45.6	1237	4	Q8NDV3	Q8ndv3 homo sapien
37	41	45.1	146	2	Q8KPT7	Q8kpt7 synechococc
38	41	45.1	242	16	P72928	P72928 synechocyst
39	41	45.1	276	16	Q8YER2	Q8yer2 listeria mo
40	41	45.1	429	10	Q80346	Q80346 raphanus sa
41	41	45.1	436	10	Q41222	Q41222 brassica ca
42	41	45.1	436	10	Q940W8	Q940w8 raphanus sa
43	41	45.1	437	10	Q39278	Q39278 brassica ca
44	41	45.1	457	12	Q8V9T2	Q8v9t2 african swi
45	41	45.1	548	12	Q64933	Q64933 avian influ

ALIGNMENTS

RESULT 1

Q9DBK8 PRELIMINARY; PRT; 902 AA.

AC Q9DBK8; 01-JUN-2001 (TREMBL) 17, Created

DT 01-JUN-2001 (TREMBL) 17, Last sequence update)

DT 01-MAR-2003 (TREMBL) 23, Last annotation update)

DE Repeat family 3 gene.

GN ITH4 OR ITRP3.

GN Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Liver;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochava H.,

RA Schirml L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Maehiro T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barch G.,

RA Blakes J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Maehima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohzuki S.,

RA Hayashizaki Y.,

RT "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:685-690(2001).

DR EMBL; AK004893; BAB23649.1; -

MED; MGI:109536; Ith4.

```
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR006587; VIT.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; vwa; 1.
DR SMART; SM00609; VIT; 1.
DR SMART; SM00327; VMA; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 902 AA; 100323 MW; 4F0BA87DA97770B CRC64;

Query Match
Best Local Similarity 55.6%; Score 57; DB 11; Length 902;
Pred. No. 0.64;
Matches 10; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KSPSEQETVLDGNLIIRY 18
Db 224 KSQSEQDPTVLNGDPTVRY 241

RESULT 2
ID 035802 PRELIMINARY; PRT; 932 AA.
AC 035802;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Inter-alpha-inhibitor H4 heavy chain.
GN ITIH4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=98153798; PubMed=9480842;
RA Soury E., Olivier E., Daveau M., Hiron M., Claessens S., Risler J.L.,
RA Sailer J.P.;
RT "The H4p heavy chain of inter-alpha-inhibitor family largely differs
RT in the structure and synthesis of its prolin-rich region from rat to
RT human";
RL Biochem. Biophys. Res. Commun. 243:522-530(1998).
DR EMBL; Y11283; CAI72155.1; -.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000387; TYR phosphatase.
DR InterPro; IPR006587; VIT.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; vwa; 1.
DR SMART; SM00609; VIT; 1.
DR SMART; SM00327; VMA; 1.
DR PROSITE; PS00383; TYR PHOSPHATASE_1; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 932 AA; 103607 MW; 6F7B172BD4F6CC48 CRC64;

Query Match
Best Local Similarity 62.6%; Score 57; DB 11; Length 932;
Pred. No. 0.66;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KSPSEQETVLDGNLIIRY 18
Db 223 KSQNEQDPTVLNGDPTVRY 240

RESULT 3
ID 091M60 PRELIMINARY; PRT; 941 AA.
AC 091M60;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Inter alpha-trypsin inhibitor, heavy chain 4.
GN ITIH4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC016500; AAH16500.1; -.
DR MGI; MGI:109536; Itih4.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR006587; VIT.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00092; vwa; 1.
DR SMART; SM00609; VIT; 1.
DR SMART; SM00327; VMA; 1.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 941 AA; 104631 MW; E7920P950A710CAF CRC64;

Query Match
Best Local Similarity 55.6%; Score 57; DB 11; Length 941;
Pred. No. 0.67;
Matches 10; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KSPSEQETVLDGNLIIRY 18
Db 224 KSQSEQDPTVLNGDPTVRY 241

RESULT 4
ID 08C7K5 PRELIMINARY; PRT; 941 AA.
AC 08C7K5;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Inter alpha-trypsin inhibitor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Liver;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK050016; BAC34032.1; -.
DR PROSITE; PS50234; VWFA; 1.
SQ SEQUENCE 941 AA; 104551 MW; DD6B70819E0731C9 CRC64;

Query Match
Best Local Similarity 62.6%; Score 57; DB 11; Length 941;
Pred. No. 0.67;
Matches 10; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KSPSEQETVLDGNLIIRY 18
Db 224 KSQSEQDPTVLNGDPTVRY 241

RESULT 5
ID 054882 PRELIMINARY; PRT; 942 AA.
AC 054882;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE PK-120.
GN ITIH4 OR ITIH-4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
```

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Balb/c; TISSUE=Liver;  
RX MEDLINE=98267140; PubMed=9602042;  
RA Cai T., Yu P., Monga S.P.S., Mishra B., Mishra L.;  
RT "Identification of mouse Itih-4 encoding a glycoprotein with two EF-hand motifs from early embryonic liver.";  
RL Biochim. Biophys. Acta 1398:32-37(1998).  
DR EMBL; AF023919; AAC25786.1; -.  
DR MGD; MGI:109536; Itih4.  
DR InterPro; IPR000087; Collagen.  
DR InterPro; IPR006587; VIT.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF00092; vwa; 1.  
DR SMART; SM00609; VIT; 1.  
DR SMART; SM00327; vwa; 1.  
DR PROSITE; PSS0234; VWF\_A; 1.  
SQ SEQUENCE 942 AA; 104668 MW; B062E1ACD4E17EAD CRC64;

Query Match 62.6%; Score 57; DB 11; Length 942;  
Best Local Similarity 55.6%; Pred. No. 0.67;  
Matches 10; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 KSPQETVLDGNLIIR 18  
DB 224 KQSEQDTVINGDPVIRY 241

## RESULT 6

Q8C7G9 PRELIMINARY; PRT; 942 AA.

AC Q8C7G9;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Inter alpha-trypsin inhibitor.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Liver;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RT the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
DR EMBL; AK050270; BAC34155.1; -.  
SQ SEQUENCE 942 AA; 104659 MW; 5D613FBA852A608 CRC64;

Query Match 62.6%; Score 57; DB 11; Length 942;  
Best Local Similarity 55.6%; Pred. No. 0.67;  
Matches 10; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 KSPQETVLDGNLIIR 18  
DB 224 KQSEQDTVINGDPVIRY 241

## RESULT 7

Q96CJ3 PRELIMINARY; PRT; 604 AA.

AC Q96CJ3;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Similar to neural cell adhesion molecule 1 (Fragment).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
NCBI\_TaxID=9606;

RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Kidney;  
RA Strauberg R.;  
RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC014205; AA014205.1; -.  
DR InterPro; IPR003961; FN\_III.  
DR InterPro; IPR007110; Ig\_Like.  
DR InterPro; IPR003598; Ig\_C2.  
DR InterPro; IPR003006; Ig\_MHC.  
DR Pfam; PF00047; Ig; 3.  
DR SMART; SM00060; FN3; 2.  
DR SMART; SM00408; IGC2; 2.  
DR PROSITE; PSS0835; IG\_LIKE; 2.  
KW Immunoglobulin domain.  
FT NON TER 1  
SQ SEQUENCE 604 AA; 66271 MW; B3B791A4D033D546 CRC64;

Query Match 52.2%; Score 47.5; DB 4; Length 604;  
Best Local Similarity 60.0%; Pred. No. 18;  
Matches 9; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

QY 3 PEQETVLDGNLIIR 17  
DB 105 PEQET-LDGHMVR 118

## RESULT 8

Q8E2L5 PRELIMINARY; PRT; 179 AA.

AC Q8E2L5;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Hypothetical protein.  
GN GB82107.  
OS Streptococcus agalactiae (serotype III).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
NCBI\_TaxID=216495;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NEM316 / Serotype III;  
RX MEDLINE=22242508; PubMed=12354221;  
RA Glaser P., Kusnuk C., Buchrieser C., Chevallier F., Frangoul L., Maeder T., Zouine M., Couve E., Lallou L., Poyart C., Tries-Cuot P., Kunst F.;  
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing invasive neonatal disease.";  
RL Mol. Microbiol. 45:1499-1513(2002).  
DR EMBL; AL766856; CAD47766.1; -.  
DR Sagalst; gbs2107; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 179 AA; 19119 MW; 7C0EB4EF4833AD56 CRC64;

Query Match 50.5%; Score 46; DB 16; Length 179;  
Best Local Similarity 52.9%; Pred. No. 8.5;  
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 SPQETVLDGNLIIR 18  
DB 147 SPQETVLDGNLIIR 163

## RESULT 9

Q8DWR6 PRELIMINARY; PRT; 179 AA.

AC Q8DWR6;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE LysM domain protein.

```

GN SAG2148.
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCB1_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=2222988; PubMed=12200547;
RA Tetselin H., Masiagnani V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Messel M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA Desoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radu D., Fedorova N.B., Scanlan D., Knouri H., Mulligan S.,
RA Catly H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Jacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaudo D., Rappolt R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014289; AAN01006.1; -.
DR TIGR; SAG2148; -.
KW Complete proteome.
SQ SEQUENCE 179 AA; 19092 MW; BPID16CF0082E4D CRC64;

Query Match 50.5%; Score 46; DB 16; Length 179;
Best Local Similarity 52.9%; Pred. No. 8.5;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 SPOQETVLDGNLIRY 18
Db 147 SPENQEKVADNMYVSRY 163

RESULT 10
O9HJZ1 PRELIMINARY; PRT; 598 AA.
AC O9HJZ1;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Hypothetical protein Ta0820.
GN TA0820.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCB1_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruopp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Mewes H.-W., Fritshman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermophilic scavenger Thermoplasma
RT acidophilum.";
RL Nature 407:508-513(2000).
DR EMBL; AL445065; CAC11949.1; -.
DR InterPro; IPR004165; COA_trans.
DR Pfam; PF01144; COA_trans; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 598 AA; 66628 MW; 3B22021E8271AA09 CRC64;

Query Match 50.5%; Score 46; DB 17; Length 598;
Best Local Similarity 61.5%; Pred. No. 32;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 4 EQOETVLDGNLIR 16
Db 483 EQEETIRGRLVI 495

RESULT 11

```

```

O9ATU5 PRELIMINARY; PRT; 319 AA.
AC O9ATU5;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE S-receptor kinase (fragment).
GN SRK.
OS Arabidopsis lyrata.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCB1_TaxID=59689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21231627; PubMed=11333247;
RA Schierup M.H., Mable B.K., Awadalla P., Charlesworth D.;
RT "Identification and characterization of a polymorphic receptor kinase
RT gene linked to the self-incompatibility locus of Arabidopsis lyrata.";
RL Genetics 158:387-399(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Schierup M.H., Mable B.K., Awadalla P., Charlesworth D.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF328999; AAK19320.2; -.
KW Receptor; Kinase.
FT NON_TER 1 1
SQ SEQUENCE 319 AA; 36043 MW; DB8F8B4D797BC5EC CRC64;

Query Match 49.5%; Score 45; DB 10; Length 319;
Best Local Similarity 44.4%; Pred. No. 24;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 1 KSPEQETVLDGNLIRY 18
Db 24 KSPVAEILANGNFVLR 41

RESULT 12
O9STE3 PRELIMINARY; PRT; 452 AA.
AC O9STE3;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Glucuronosyl transferase-like protein.
GN T6H20.280.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OX eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCB1_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Schoiane N., Robert C., Brottier P., Wincker P., Catolico L.,
RA Aricquenave F., Saurin W., Weissenbach J., Mewes H.-W., Mayer K.F.X.,
RA Lemcke K., Schueller C., Quetier F., Salanoubat M.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL096859; CAB51196.1; -.
DR InterPro; IPR002213; UDP_gluco_trans.
DR Pfam; PF00201; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 452 AA; 50525 MW; E0F29354F4284A63 CRC64;

Query Match 49.5%; Score 45; DB 10; Length 452;
Best Local Similarity 50.0%; Pred. No. 35;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

```

QY 1 KSPQOETVLDGNLIRY 18  
DB 160 KDPKQDKVLEGLPLRY 177

RESULT 13

08GZ65 PRELIMINARY; PRT; 452 AA.  
AC 08GZ65;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DE Putative glucuronosyl transferase.  
GN A13G4690/T6H20.280.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Saki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,  
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,  
RA Hayashizaki Y., Shinzaki K.,  
RT "Arabidopsis thaliana full-length cDNA."  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
KW EMBL: AK117184; BAC1861.1; -.  
SQ SEQUENCE 452 AA; 50524 MW; BEFC9354F422EA63 CRC64;

Query Match 49.5%; Score 45; DB 10; Length 452;  
Best Local Similarity 50.0%; Pred. No. 35;  
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 KSPQOETVLDGNLIRY 18  
DB 160 KDPKQDKVLEGLPLRY 177

RESULT 14

08BV52 PRELIMINARY; PRT; 1113 AA.  
AC 08BV52;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DE Protein-tyrosine phosphatase.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Thymus;  
RX MEDLINE=2354683; PubMed=12466851;  
RA The RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
DR EMBL: AK080313; BAC37876.1; -  
SQ SEQUENCE 1113 AA; 124999 MW; 8558E42299303C71 CRC64;

Query Match 49.5%; Score 45; DB 11; Length 1113;  
Best Local Similarity 69.2%; Pred. No. 92;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 PEOQETVLDGNLI 15  
DB 551 PROJETSLGNNLI 563

RESULT 15  
Q28006 PRELIMINARY; PRT; 2484 AA.  
AC Q28006;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DE B414 tyrosine phosphatase (EC 3.1.3.48).  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Vega O.C., Walton K.M., Dixon J.E.,  
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O = PROTEIN  
CC TYROSINE + PHOSPHATE.  
CC -1- SIMILARITY: CONTAINS A DOMAIN FOUND IN BAND 4.1, EZRIN, MOESIN,  
CC RADIXIN, AND TALIN.  
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-  
CC TYROSINE PHOSPHATASE FAMILY.  
DR EMBL: U20807; AAA73516.1; -.  
DR HSP; Q12923; 3PDZ.  
DR InterPro: IPR000299; Band\_4.1.  
DR InterPro: IPR001478; PDZ.  
DR InterPro: IPR000387; Tyr\_phosphatase.  
DR InterPro: IPR000242; Tyr\_PP.  
DR Pfam: PF00373; Band\_41; 1.  
DR Pfam: PF00595; PDZ; 5.  
DR Pfam: PF00102; Y\_phosphatase; 1.  
DR PRINTS; PR00935; BAND41.  
DR PRINTS; PR00700; PRTYPHPTASE.  
DR SMART; SM00295; B41; 1.  
DR SMART; SM00228; PDZ; 5.  
DR SMART; SM00194; PTPC; 1.  
DR PROSITE; PS00660; BAND\_41.1; FALSE\_NEG.  
DR PROSITE; PS00661; BAND\_41.2; FALSE\_NEG.  
DR PROSITE; PS00507; BAND\_41\_3; 1.  
DR PROSITE; PS50106; PDZ; 5.  
DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; FALSE\_NEG.  
DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PP; 1.  
KW Structural protein; Cytoskeleton; Hydrolase; Coiled coil.  
FT DOMAIN 600 800 BAND 4.1-LIKE DOMAIN.  
FT DOMAIN 1341 1344 POLY-SER.  
FT DOMAIN 2236 2484 PROTEIN-TYROSINE PHOSPHATASE.  
FT ACT SITE 2407 2407 BY SIMILARITY.  
FT DOMAIN 467 504 COILED COIL (POTENTIAL).  
FT DOMAIN 1761 1793 COILED COIL (POTENTIAL).  
SQ SEQUENCE 2484 AA; 276382 MW; 45A92F0DAF1ED13D CRC64;

Query Match 49.5%; Score 45; DB 6; Length 2484;  
Best Local Similarity 69.2%; Pred. No. 2,2e+02;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 PEOQETVLDGNLI 15  
DB 462 PROJETSLGNNLI 474

RESULT 16  
Q39363 PRELIMINARY; PRT; 409 AA.  
AC Q39363;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE S-1 locus glycoprotein precursor.  
GN SLG.  
OS Brassica napus (Rape).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;





DR Pfam; PF00954; S locus glycop; 1.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR SMART; SM00108; B lectin; 1.  
 DR SMART; SM00473; PAN Ap; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE\_ATP; 1.  
 DR PROSITE; PS0011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 SO SEQUENCE 856 AA; 97372 MW; FE601B2E5E9D72E5 CRC64;

Query Match 48.4%; Score 44; DB 10; Length 856;  
 Best Local Similarity 44.4%; Pred. No. 1e+02;  
 Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 1 KSPQETVLDGNLIIRY 18  
 Db 130 RSPVAVELANGNVIRY 147

RESULT 20  
 ID 08VR20 PRELIMINARY; PRT; 401 AA.  
 AC 08VR20;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE 60 kDa heat shock protein groEL (60 kDa chaperonin) (Protein Cpn60)  
 DE (groEL protein) (Fragment).  
 GN GROEL.  
 OS Chlamydomophila felis.  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydomophila.  
 OX NCBI\_TaxId=83556;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FEIS;  
 RA Di Francesco A., Battilani M., Baldeili R.;  
 RT "PCR amplification and sequencing of groEL homolog gene of  
 RT Chlamydomophila felis."  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND  
 CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS  
 CC CONDITIONS (BY SIMILARITY).  
 CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF  
 CC 7 SUBUNITS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.

DR EMBL; AF448139; AAJ38954.1; -  
 DR InterPro; IPR001844; Chaperin\_Cpn60.  
 DR InterPro; IPR002423; Cpn60/TCP-1.  
 DR Pfam; PF00118; Cpn60 TCP1; 1.  
 DR PRINTS; PR00304; TCOMPLEXTCP1.  
 DR PROSITE; PS00298; CHAPERONIN60.  
 DR PROSITE; PS00296; CHAPERONIN\_CPN60; 1.  
 DR ATP-binding; Chaperone; Heat shock.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SO SEQUENCE 401 AA; 43264 MW; A06E0B8FCEA440AA CRC64;

Query Match 47.3%; Score 43; DB 2; Length 401;  
 Best Local Similarity 47.1%; Pred. No. 67;  
 Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 2 SPQOETVLDGNLIIRY 18  
 Db 150 NPETOCCVLESVLVLY 166

RESULT 21  
 ID 08NMF5 PRELIMINARY; PRT; 628 AA.  
 AC 08NMF5;  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)

DE Hypothetical protein MW1443.  
 GN MW1443.  
 OS Staphylococcus aureus (strain MW2).  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxId=196620;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22040717; PubMed=12044378;  
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
 RA Nagai Y., Iwama N., Asano K., Naito T., Kuroda H., Chi L.,  
 RA Yamamoto K., Hiratake K.;  
 RT "Genome and virulence determinants of high virulence community-  
 RT acquired MRSA."  
 RL Lancet 359:1819-1827(2002).  
 DR EMBL; AP004827; BAB95308.1; -  
 KM Hypothetical protein, Complete proteome.  
 SO SEQUENCE 628 AA; 71562 MW; 40F7C427B986B4D CRC64;

Query Match 47.3%; Score 43; DB 16; Length 628;  
 Best Local Similarity 53.3%; Pred. No. 1.1e+02;  
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 4 BQOETVLDGNLIIRY 18  
 Db 450 QODATIDGNVQINRY 464

RESULT 22  
 ID 094GY4 PRELIMINARY; PRT; 736 AA.  
 AC 094GY4;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Putative snRNP protein.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriactoidae; Oryzae; Oryza.  
 OX NCBI\_TaxId=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Buell C.R., Yuan Q., Ouyang S., Moffat K.S., Hill J.N., Gansberger K.,  
 RA Brenner M., Burgess S., Hance M., Shvartsbeyn M., Telitin T.,  
 RA Riggs F., Heiao J., Ziemann V., Blunt S., Pal G., VanAken S.E.,  
 RA Uteback T.R., Feldblum T.V., Quackenbush J., Salzberg S.L.,  
 RA White O., Fraser C.M.;  
 RT "Oryza sativa chromosome 3 BAC OSJNBa0018H01 genomic sequence."  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.

DR EMBL; AC087181; AAK8497.1; -  
 DR Gramene; Q94GY4; -  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR000629; DEAD\_BOX.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR Pfam; PF00270; DEAD; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELICG; 1.  
 DR PROSITE; PS00039; DEAD\_ATP\_HELICASE; 1.  
 DR ATP-binding; Helicase; Helicase; Hydrolase.  
 SO SEQUENCE 736 AA; 84520 MW; 5D7ED635FC1AD629 CRC64;

Query Match 47.3%; Score 43; DB 10; Length 736;  
 Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 KSPQOETVLDG 12  
 Db 612 KSPQOETVLDG 623

RESULT 23  
 ID Q8EL73 PRELIMINARY; PRT; 749 AA.  
 AC Q8EL73;  
 DT 01-MAR-2003 (TRENBLrel. 23, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DE DNA mismatch repair protein Hexa.  
 GN Hexa OR SAG2101.  
 OS Streptococcus agalactiae (serotype V).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC NCBI\_TaxID=216466;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H1E831 / DSM 14371 / JCM 11309;  
 RX MEDLINE=22220767; PubMed=1235376;  
 RA Takami H., Takaki Y., Uchiyama I.;  
 RT "Genome sequence of Oceanobacillus theyensis isolated from the Iheya  
 RT Ridge and its unexpected adaptive capabilities to extreme  
 RT environments.";  
 RL Nucleic Acids Res. 30:3927-3935 (2002).  
 DR EMBL; AF004604; BAC15314.1; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 749 AA; 86306 MW; 9495B08BFF246DFA CRC64;

Query Match 47.3%; Score 43; DB 16; Length 749;  
 Best Local Similarity 41.2%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Oy 2 SPOQETVLDGNLI 18  
 Db 381 SPVNDETIKDGHILFSW 397

RESULT 24  
 ID Q8E2R3 PRELIMINARY; PRT; 858 AA.  
 AC Q8E2R3;  
 DT 01-MAR-2003 (TRENBLrel. 23, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DE DNA mismatch repair protein Muts.  
 GN Muts OR GBS2054.  
 OS Streptococcus agalactiae (serotype III).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC NCBI\_TaxID=216495;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NEW316 / Serotype III;  
 RX MEDLINE=22242508; PubMed=12354221;  
 RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,  
 RA Masdek T., Zouine M., Couve E., Lalioui L., Poyart C., Tieu-Cuot P.,  
 RA Kunst P.;  
 RT "Genome sequence of Streptococcus agalactiae, a pathogen causing  
 RT invasive neonatal disease.";  
 RL Mol. Microbiol. 45:1499-1513 (2002).  
 DR EMBL; AL766856; CAD47713.1; -  
 KW Sagalact; gbs2054; -  
 SO SEQUENCE 858 AA; 97078 MW; EB7FB2EB0153BESF CRC64;

Query Match 47.3%; Score 43; DB 16; Length 858;  
 Best Local Similarity 50.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Oy 2 SPOQETVLDGNLI 15  
 Db 400 APEAQATITEGNI 413

RESULT 25  
 Q8DWM1

ID Q8DWM1 PRELIMINARY; PRT; 858 AA.  
 AC Q8DWM1;  
 DT 01-MAR-2003 (TRENBLrel. 23, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DE DNA mismatch repair protein Hexa.  
 GN Hexa OR SAG2101.  
 OS Streptococcus agalactiae (serotype V).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC NCBI\_TaxID=216466;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=2603 V/R / Serotype V;  
 RX MEDLINE=22222988; PubMed=1200547;  
 RA Tettelein H., Masignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,  
 RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,  
 RA Macleod L.C., Wolf A.M., Beaman M.J., Brinkac L.M., Dougherty S.C.,  
 RA Debay R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,  
 RA Radu D., Fedorova N.B., Scanlan D., Khouri H., Mulligan S., Mora M.,  
 RA Jacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,  
 RA Rieraudo D., Rapunoli R., Telford J.L., Kasper D.L., Grandi G.,  
 RA Fraser C.M.;  
 RT "Complete genome sequence and comparative genomic analysis of an  
 RT emerging human pathogen, serotype V Streptococcus agalactiae.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).  
 DR EMBL; AE014286; AAN00959.1; -  
 KW TIGR; SAG2101; -  
 SQ SEQUENCE 858 AA; 97078 MW; EB7FB2EB0153BESF CRC64;

Query Match 47.3%; Score 43; DB 16; Length 858;  
 Best Local Similarity 50.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Oy 2 SPOQETVLDGNLI 15  
 Db 400 APEAQATITEGNI 413

RESULT 26  
 ID Q9ATU8 PRELIMINARY; PRT; 312 AA.  
 AC Q9ATU8;  
 DT 01-JUN-2001 (TRENBLrel. 17, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DE S-receptor kinase (Fragment).  
 GN SRK.  
 OS Arabidopsis lyrata.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=59689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21231627; PubMed=11333247;  
 RA Schierup M.H., Mable B.K., Awadalla P., Charlesworth D.;  
 RT "Identification and characterization of a polymorphic receptor kinase  
 RT gene linked to the self-incompatibility locus of Arabidopsis lyrata.";  
 RL Genetics 158:387-399 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Schierup M.H., Mable B.K., Awadalla P., Charlesworth D.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF328996; AAK19317.2; -  
 KW Receptor; Kinase.  
 FT NON TER 1  
 FT 312 312  
 SQ SEQUENCE 312 AA; 35680 MW; 70DA31CF6F952A7 CRC64;

Query Match 46.2%; Score 42; DB 10; Length 312;

Best Local Similarity 50.0%; Pred. No. 76;  
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
QY 1 KSPQOETVLDGNLIIRY 18  
Db 23 RSPVVALDLDTGNLVIRY 40

RESULT 27  
Q9EUS PRELIMINARY; PRT; 360 AA.

AC Q9EUS; 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
OS Salmomella enterica subsp. enterica serovar Choleraesuis.  
OC Plasmid 50k virulence plasmid.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmomella.  
OX NCBI\_Taxid=119912;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RF-1;  
RX MEDLINE=21153631; PubMed=11254626;  
RA Hamed T., Okada N., Nakazawa N., Kawakami T., Danbara H.;  
RT "Complete DNA Sequence and Comparative Analysis of the 50-Kilobase  
Virulence Plasmid of Salmomella enterica Serovar Choleraesuis.";  
RU Infect. Immun. 69:2612-2620 (2001).  
DR EMBL; AB040415; BAB20537.1; -.  
KM Plasmid.  
SQ SEQUENCE 360 AA; 39516 MW; 4DC216B053940AA7 CRC64;  
QY Query Match 46.2%; Score 42; DB 2; Length 360;  
Best Local Similarity 66.7%; Pred. No. 89;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
Db 137 QSEIVTLGNLVM 148

RESULT 28  
023853 PRELIMINARY; PRT; 428 AA.  
ID 023853;  
AC 023853; 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
OS S glycoprotein (Fragment).  
OC Brassica campestris (Field mustard).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OX NCBI\_Taxid=3711;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97352858; PubMed=9207151;  
RA Kusaba M., Nishio T., Satoh Y., Hinata K., Ockendon D.;  
RT "Striking sequence similarity in inter- and intra-specific comparisons  
of a class I SIG alleles from Brassica oleracea and Brassica  
campestris: Implications for the evolution and recognition  
mechanism.";  
RU Proc. Natl. Acad. Sci. U.S.A. 94:7673-7678 (1997).  
RL EMBL; D85219; BAA21953.1; -.  
DR InterPro; IPR001480; B\_lectin.  
DR InterPro; IPR003609; Pan app.  
DE Pfam; PF01453; Stocus glycop.  
DR Pfam; PF00954; S\_locus\_glycop. 1.  
DR SMART; SM00108; B\_lectin; 1.  
DR SMART; SM00473; PAN\_Ap; 1.  
FT NON\_TER 1

FT NON\_TER 428 428  
SQ SEQUENCE 428 AA; 48738 MW; D53C5A2CBDD34D53 CRC64;  
Query Match 46.2%; Score 42; DB 10; Length 428;  
Best Local Similarity 44.4%; Pred. No. 1.1e+02;  
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;  
QY 1 KSPQOETVLDGNLIIRY 18  
Db 120 RSPVVALDLDTGNLVIRY 137

RESULT 29  
Q9S6B0 PRELIMINARY; PRT; 497 AA.

AC Q9S6B0; 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DE 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE 60 kD heat shock protein GroEL (60 kDa chaperonin) (Protein Cpn60)  
DE (GroEL protein) (Fragment).  
GN GROEL.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_Taxid=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AR388;  
RA Kaldenbeck B., Gao D.;  
RT "PCR amplification and sequencing of the partial coding region of  
homologs of Chlamydia of the 60 kDa heat shock protein GroEL.";  
RU Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND  
PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS  
CONDITIONS (BY SIMILARITY).  
CC -1- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF  
7 SUBUNITS (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.  
DR EMBL; AF109791; AAD26145.1; -.  
DR HSP; P06139; IGRL.  
DR InterPro; IPR001844; Chaperin Cpn60.  
DR InterPro; IPR002423; Cpn60/TCP-1.  
DR Pfam; PF00118; Cpn60 TCP1. 1.  
DR PRINTS; PR00298; CHAPERONIN60.  
DR PRINTS; PR00304; TCOMPLEXTCP1.  
DR PROSITE; PS00296; CHAPERONIN\_CPN60. 1.  
KW ATP-binding; Chaperone; Heat shock.  
FT NON\_TER 1  
FT NON\_TER 497  
SQ SEQUENCE 497 AA; 53101 MW; 5DF6B064F9D5F65 CRC64;  
QY Query Match 46.2%; Score 42; DB 2; Length 497;  
Best Local Similarity 52.9%; Pred. No. 1.3e+02;  
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
Db 185 NPFTQECVLDALILY 201

RESULT 30  
001963 PRELIMINARY; PRT; 858 AA.  
ID 001963;  
AC 001963; 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Serine/threonine kinase receptor precursor (EC 2.7.1.37) (S-receptor  
kinase) (SRK) (S-locus receptor kinase).  
GN SRK-910.  
OS Brassica napus (Rape).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

CC eurosida II; Brassicales; Brassicaceae; Brassica.  
 OX NCB1\_TaxID=3708;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. OLEIFERA; TISSUE=PISTIL;  
 RX MEDLINE=93076111; PubMed=1332796;  
 RA Goring D.R., Rothstein S.J.;  
 RT "The S-locus receptor kinase gene in a self-incompatible Brassica  
 napus line encodes a functional serine/threonine kinase.";  
 RL Plant Cell 4:1273-1281 (1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. M1;  
 RA Cui Y., Brugliere N., Jackman L., Bi Y.M., Rothstein S.J.;  
 RT "Structural and transcriptional comparative analysis of the S-locus  
 regions in two self-incompatible Brassica napus lines.";  
 RL Plant Cell 11:2217-2231 (2000).  
 CC -1- FUNCTION: INVOLVED IN SPOROPHYTIC SELF-INCOMPATIBILITY SYSTEM (THE  
 INABILITY OF FLOWERING PLANTS TO ACHIEVE SELF-FERTILIZATION),  
 PROBABLY ACTING IN COMBINATION WITH S-LOCUS-SPECIFIC  
 GLYCOPROTEINS. INTERACTION WITH A LIGAND IN THE CYTOPLASMIC  
 DOMAIN TRIGGERS THE PROTEIN KINASE ACTIVITY OF THE CYTOPLASMIC  
 DOMAIN.  
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN = ADP + A PHOSPHOPROTEIN.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY IN THE PISTIL AND TO A MUCH  
 LESSER EXTENT, IN THE ANTHOR.  
 CC -1- MISCELLANEOUS: THERE ARE A NUMBER OF DIFFERENT S ALLELES IN  
 B. NAPUS. POSSIBLY PROVIDING THE RECOGNITION SPECIFICITY.  
 CC -1- SIMILARITY: THE EXTRACELLULAR DOMAIN IS SIMILAR TO S-LOCUS  
 GLYCOPROTEINS OF BRASSICA, WHILE THE INTRACELLULAR DOMAIN IS A  
 SER/THR-PROTEIN KINASE RELATED TO RAF KINASES.  
 CC EMBL; M97667; AAA33008.1; -.  
 DR EMBL; AJ245479; CAB89179.1; -.  
 DR InterPro; IPR001480; B\_lectin.  
 DR InterPro; IPR003609; Pan\_app.  
 DR InterPro; IPR000719; Prot\_Kinase.  
 DR InterPro; IPR002290; Ser\_thr\_kinase.  
 DR InterPro; IPR000858; Slocus\_glycop.  
 DR Pfam; PF01453; Agglutinin; 1.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR Pfam; PF00954; S\_locus\_glycop; 1.  
 DR ProDom; PD000001; Prot\_Kinase; 1.  
 DR SMART; SM00108; B\_lectin; 1.  
 DR SMART; SM00473; PAN\_AP; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW Transferrase; Serine/threonine-protein kinase; signal; ATP-binding;  
 KW Transmembrane; Receptor; Glycoprotein; Self-incompatibility;  
 KW Polymorphism.  
 FT SIGNAL 1 31 POTENTIAL.  
 FT CHAIN 32 858 S RECEPTOR KINASE (POTENTIAL).  
 FT DOMAIN 32 447 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 448 467 POTENTIAL.  
 FT DOMAIN 468 858 CYTOPLASMIC (POTENTIAL).  
 FT NP\_BIND 535 543 ATP (BY SIMILARITY).  
 FT ACT\_SITE 654 654 ATP (BY SIMILARITY).  
 FT BINDING 557 557 ATP (BY SIMILARITY).  
 FT CARBOHYD 48 48 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 158 158 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 393 393 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 858 AA; 97908 MW; 675F8B4B894EA3F7 CRC64;

Query Match 46.2%; Score 42; DB 10; Length 858;  
 Best Local Similarity 44.4%; Pred. No. 2.3e+02;  
 Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 1 KSPQQETVLQGNLIIIRY 18

DB 130 RSPVABELENGENFVIRY 147  
 :|| :|| :||  
 RESULT 31  
 ID 093GL4 PRELIMINARY; PRT; 1752 AA.  
 AC 093GL4  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Conjugative transfer: oriT nicking-unwinding.  
 GN TRAI OR PSLT108.  
 OS Salmonella typhimurium.  
 CC Plasmid pSLT.  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Salmonella.  
 OX NCB1\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Lrt2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Muiyaney E.,  
 RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 Lrt2.";  
 RL Nature 413:852-856 (2001).  
 DR EMBL; AE006471; AL23509.1; -.  
 DR InterPro; IPR001452; SH3.  
 DR SMART; SM00326; SH3; 1.  
 KW Plasmid; Complete proteome.  
 SQ SEQUENCE 1752 AA; 191683 MW; 6182AC451D5B81F8 CRC64;

Query Match 46.2%; Score 42; DB 16; Length 1752;  
 Best Local Similarity 66.7%; Pred. No. 5e+02;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 5 QOETVLQGNLII 16  
 |||||  
 DB 130 QSETVLGTNLVW 141

RESULT 32  
 ID 09WTB0 PRELIMINARY; PRT; 1756 AA.  
 AC 09WTB0  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE DNA helicase I.  
 GN TRAI.  
 OS Escherichia coli.  
 CC Plasmid R100.  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Escherichia.  
 OX NCB1\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sampe G., Mizobuchi K.;  
 RT "Organization and diversification of plasmid genomes: complete  
 nucleotide sequence of the R100 genome.";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP000342; BAA7885.1; -.  
 KW Helicase.  
 SQ SEQUENCE 1756 AA; 191683 MW; 56477ED2DC915BB3 CRC64;

Query Match 46.2%; Score 42; DB 2; Length 1756;  
 Best Local Similarity 66.7%; Pred. No. 5e+02;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 5 OQETVLDSNLI 16  
DB 130 QSETVLGNLVM 141

## RESULT 33

O9W2J8 PRELIMINARY; PRT; 2515 AA.

ID O9W2J8  
AC O9W2J8;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
DE TUD protein.  
GN TUD OR CG9450.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
NCBI\_TaxID=7227;  
[1]  
SEQUENCE FROM N.A.  
RP STRAIN=BERKELEY;  
RC MEDLINE=20196006; PubMed=10731132;  
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Amaraides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
Brandon R.C., Rogers Y.-H.C., Blazer G., Chame M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abail J.F., Agbayant A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
Balew R.M., Baau P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,  
Buttis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Paolo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
Hosbin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kerchum K.A.,  
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,  
Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
Shue B.C., Sider-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
Svirbas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wasserman D.T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.W., Venter J.C.;  
"The genome sequence of Drosophila melanogaster";  
Science 287:2185-2195(2000).  
EMBL; AB003453; AAF46693.1; -.  
HSSP; Q16637; IG5V.  
DR FlyBase; FBgn003891; tud.  
DR InterPro; IPR001097; Maternal\_tudor.  
DR InterPro; IPR002999; Tudor.  
DR Pfam; PF00567; tudor, 10.  
DR SMART; SMO0333; tudor, 9.  
DR PROSITE; PSS0304; tudor, 9.  
SEQUENCE 2515 AA; 285264 MW; 1218700174D66701 CRC64;

Query Match 46.2%; Score 42; DB 5; Length 2515;  
Best Local Similarity 42.9%; Pred. No. 7.3e+02;

Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;  
Oy 1 KSPQETVLDSNLI 14  
DB 959 QKFOROKSTLDGNT 972

## RESULT 34

O8CQAI PRELIMINARY; PRT; 346 AA.

ID O8CQAI  
AC O8CQAI;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Branched-chain alpha-keto acid dehydrogenase E1.  
GN SE0255.  
OS Staphylococcus epidermidis.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OC NCBI\_TaxID=1282;  
[1]  
RP STRAIN=ATCC 12228;  
RC Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,  
Chen Z., Wen Y.;  
Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AB016744; AAC03852.1; -.  
KW Complete proteome.  
SO SEQUENCE 346 AA; 36907 MW; D86017CE4A66E26A CRC64;

Query Match 45.6%; Score 41.5; DB 16; Length 346;  
Best Local Similarity 55.0%; Pred. No. 1e+02;  
Matches 11; Conservative 1; Mismatches 3; Indels 5; Gaps 1;

Oy 2 SPQOETVLDSNLI 16  
DB 254 SPWQOETVLDSVKKTGRLIV 273

## RESULT 35

O9Y3G5 PRELIMINARY; PRT; 806 AA.

ID O9Y3G5  
AC O9Y3G5;  
DT 01-NOV-1999 (TREMBlrel. 12, Created)  
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE D102D24.1 (Novel mitosis-specific chromosome segregation protein SMCI  
like protein) (Fragment).  
GN BK268H5.5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OC NCBI\_TaxID=9606;  
[1]  
RP SEQUENCE FROM N.A.  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL021911; CAB41703.1; -.  
DR InterPro; IPR003395; SMC\_N.  
DR Pfam; PF02463; SMC\_N; 1.  
FT NON TER 806  
SEQUENCE 806 AA; 94458 MW; 2063C5E4F1841B99 CRC64;

Query Match 45.6%; Score 41.5; DB 4; Length 806;  
Best Local Similarity 56.2%; Pred. No. 2.6e+02;  
Matches 9; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

Oy 3 PEOQETVLDSNLI 17  
DB 633 PERQKTVLDSNLI 648

## RESULT 36

O8NDV3

```

ID Q8NDV3 PRELIMINARY; PRT; 1237 AA.
AC Q8NDV3;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE SMC1beta protein.
GN SMC1BETA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Cobbe N., Heck M.M.S.;
RT "Phylogenetic analysis of SMC proteins.";
RL Submitted (AUG-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ504806; CAD3404.1; -
DR Genew; HGNC:11112; SMC12.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02483; SMC_C; 1.
DR Pfam; PF02463; SMC_N; 1.
SQ SEQUENCE 1237 AA; 144192 MW; 801ADA16F4474330 CRC64;

Query Match 45.1%; Score 41; DB 4; Length 1237;
Best Local Similarity 56.2%; Pred. No. 4.1e+02;
Matches 9; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

OY 3 PEOGETVLDGNLIR 17
Db 633 PERQTVALDGTFLK 648

RESULT 37
O8KPT7 PRELIMINARY; PRT; 146 AA.
AC O8KPT7;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical protein.
GN SED0021.
OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=PCC 7492;
RC Holtman C.K., Sandoval P., Chen Y., Socias T., Mohler B.J.,
RA Gonzalez A., Salinas I., McMurtry S., Golden S.S., Youderian P.;
RT "Synechococcus elongatus PCC7942 cosmid 6C3.";
RL Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY10852; AAM82654.1; -
KW Hypothetical protein.
SQ SEQUENCE 146 AA; 16549 MW; FE1C4AD8B3FC976 CRC64;

Query Match 45.1%; Score 41; DB 2; Length 146;
Best Local Similarity 52.9%; Pred. No. 50;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 SPEOETVLDGNLIR 18
Db 67 SPRGETVVGRLKRY 83

RESULT 38
P72928 PRELIMINARY; PRT; 242 AA.
AC P72928;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)

```

```

DE Hypothetical protein s111022.
GN S111022.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,
RA Hosonouchi T., Matsuno A., Muraki A., Nakazaki N., Natsu K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yanada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90901; BA16945.1; -
DR InterPro; IPR005496; Terc.
DR Pfam; PF03741; Terc; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 242 AA; 26782 MW; CB388B39EB02C184 CRC64;

Query Match 45.1%; Score 41; DB 16; Length 242;
Best Local Similarity 41.2%; Pred. No. 86;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 2 SPEOETVLDGNLIR 18
Db 51 NPDQRRALNAGVIAT 67

RESULT 39
O8Y6R2 PRELIMINARY; PRT; 276 AA.
AC O8Y6R2;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein lmo1622.
GN LMO1622.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=EGD-e / Serovar 1/2a;
RC MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Bachthieser C., Ruenick C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Cherouni F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Ficht H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kutaprat H.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Varquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591980; CAC99700.1; -
DR ListList; LMO01622; -
DR InterPro; IPR000631; UPF0031.
DR Pfam; PF01256; UPF0031; 1.
DR TIGRfams; TIGR00196; YJef_cterm; 1.
DR PROSITE; PS01050; UPF0031_2; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 276 AA; 29796 MW; 0EAFD71DEC7157F CRC64;

Query Match 45.1%; Score 41; DB 16; Length 276;
Best Local Similarity 44.4%; Pred. No. 99;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

```

Qy 1 KSPQOETVLDGNLIRY 18  
 ||| : : : : :  
 Db 118 KSTERQOVITDGDGITTY 135

## RESULT 40

080346 PRELIMINARY; PRT; 429 AA.  
 ID 080346  
 AC 080346;  
 DT 01-NOV-1998 (TRENBLrel. 08, Created)  
 DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
 DE S glycoprotein (Fragment).  
 GN SLG(S1).  
 OS Raphanus sativus (Rash).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Rapanus.  
 OC NCBI\_TaxID=3726;  
 CX [1]  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98311079; PubMed=9648745;  
 RA Sakamoto K., Kusaba M., Nishio T.;  
 RT "Polymorphism of the S-locus glycoprotein gene (SLG) and the S-locus  
 RT related gene (SLR1) in Raphanus sativus L. and self-incompatible  
 RT ornamental plants in the Brassicaceae."  
 RL Mol. Gen. Genet. 258:397-403(1998).  
 DR EMBL; AB009677; BAA31724.1; -  
 DR InterPro; IPR001480; B\_lectin.  
 DR InterPro; IPR003609; Pan\_app.  
 DR InterPro; IPR000858; Slocus\_glycop.  
 DR Pfam; PF01453; Agglutinin; 1.  
 DR Pfam; PF00954; S-locus\_glycop; 1.  
 DR SMART; SM00108; B\_lectin; 1.  
 DR SMART; SM00473; PAN\_AP; 1.  
 FT NON\_TER 1  
 FT NON\_TER 429  
 FT SEQUENCE 429 AA; 48543 MW; 0C53A10662DDF938 CRC64;

Query Match 45.1%; Score 41; DB 10; Length 429;  
 Best Local Similarity 38.9%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
 Qy 1 KSPQOETVLDGNLIRY 18  
 ||| : : : : :  
 Db 120 RSPVVALIANGNFVRY 137

Search completed: October 10, 2003, 17:24:43  
 Job time : 97 secs

**This Page Blank (uspto)**